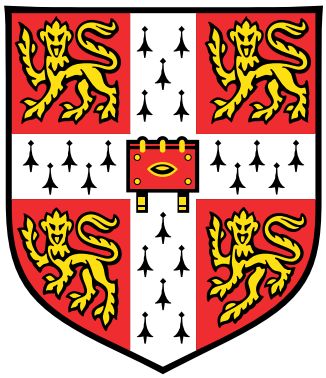
Exotic vortices in superfluids and matrix product states for quantum optimization and machine learning



Beñat Mencia Uranga

Gonville & Caius College
University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

To Aita, Ama, Uma, Aitite and the Amamas.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 60,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Chapters 2 and 4 of the thesis are based on research published in:

- B. Mencia Uranga and Austen Lamacraft, "*Infinite lattices of vortex molecules in Rabi-coupled condensates*", Phys. Rev. A 97, 043609 (2018)
- Beñat Mencia Uranga and Austen Lamacraft, "*SchrödingeRNN: Generative Modeling of Raw Audio as a Continuously Observed Quantum State*", arXiv:1911.11879

Beñat Mencia Uranga
July 2019

Acknowledgments

First and foremost, I thank my supervisor Austen Lamacraft. My admiration for him made me try to absorb as much as I could from his knowledge, style and taste. In the early days, when I asked him what subfield in condensed matter physics he belonged to, he told me about *Birds and Frogs* by Freeman Dyson. He said: *'We are trying to be birds'*. He taught me that there is a value in being open to work on new problems, without being too conservative or fearful. From him I learnt the concept of *mathematical beauty*, I could always smell it when working with him ... He supported me when I decided that I wanted to switch gears and get immersed in the field of machine learning. The conversation we had on prospective career paths, which I still remember with a great level of detail, was crucial for me to decide to go down the non-academic AI research path. I cannot thank him enough for this conversation. It is due to what I have learned from working with him that I now feel confident to fly the *bird* nest and glide through the sky of Silicon Valley.

Secondly, I would like to thank Attila and Ollie (< 3) for all the hours they devoted to proofread my thesis. The level of detail of their comments was incredible, both regarding style and content. I also thank them for the super detailed analysis of my cover letter that got me my dream job. Szymosiu also helped with the proofreading.

I thank Kasia for many fruitful discussions and for helping to design our *quantum-inspired* model.

I thank *The office*, my good friends: Adam-*Padre*, Daniel-*comilón*, Chris-*Parmesan* and Simon-*Szymmosiu*. They made the journey a pleasant journey. I celebrated every morning that I had to go to the office. They taught me a lot of physics. They taught me English and it is thank to them that I have the posh British accent I have (smiley). I thank them for being lovely people.

I thank Michael Rutter and Gareth Conduit. Michael is devoted to make TCM people's life easy. I learnt a lot about computers and numerical precision from him. I will always remember the road trip with Gareth to Poland: *hot chick project*. He taught me about career strategy. They are the everlasting couple of TCM.

I thank Emilio Artacho, he has been a mentor to me. He planted the seed of coming to Cambridge. To a large degree, it is due to him that I did my PhD in TCM.

I thank Dima Khmelnitskii for being the sweet-grumpy grandpa of TCM. For the amazing *Fairy tails* and for the old-school Russian physics he taught us.

I thank the rest of TCM friends and colleagues: Bruninho, Salvatore-*Mufasa*, Aurelio, Marianne, Daniel Rowlands, Bartek, Victor, Philipp, Ben Irwin-*Gong xí*, Chau-*shan*, Ezequiel-*gato*, Joas-*comilón*, Stephen, Yang, James, Petr-*pal*, Tycho-*Santiago*, Max, Angela, Darryl, Nick, Proffessor Wade (*or teacher if you will*), Nicoló, Mateo-*fratello*, Ivona-*chica*, Benji-*Remez*, Tobias-*Schaich*, Jan-*Janny boy*, Andrea-*Pizzi*, Elis, Raul-*txutxa hueon* ... and senior members: Nigel, Claudio, Richard, Tomeu, Andreas, Johannes, Dima Kovrizhin ... They made TCM a pleasant arena to play with science, a factory of desire of knowledge. It produces a remarkable feeling to be surrounded by brilliant minds without it making it a henhouse of many roosters, but a friendly homy family environment.

But Cambridge was *a lot more* than physics and TCM. Gonville & Caius College has been my home and their peoples my family.

The gang on first year: Artane-*Artichoke*, Jacopo-*Pistorial*, Daniel-*Alleman* and David Lilienfeld. All the *cheaky Fifas* and the hours of music and philosophy in the basements of Harvey Road. The chocolate mousse and the food from the other side of the bridge.

Julio-*kakati* and Wiktor-*how do you think?* I thank them for all the fun we had together. The *magic house* with Santiago & Co. Our house was the *Renaissance house*, the house of the peoples and the brunches.

A special mention goes to Russel Moore-*Feynman*, a brother to me. I thank him for all the great nights of *pub philosophy*.

The *Caius Movie Appreciation Society* has been one of the main sources of joy throughout my PhD. I thank Michal-*The Stag*, Daniela, Alex, Guo, Agustin & Gloria, Michele and many other people that have been part of it.

I am grateful to *Caius Boat Club*, rowing was a lot of fun! Four years of rowing, three generations of the Caius MCR boat, seven Bumps campaigns and four blades with: Ben Trigg, George Tetley, David Baumgärtner, Joep, Hugo-*skux*, Chris-*the chopper*, Lander, Hugo Larose, Jiles, Alessandro, Anthony, Mathew Holland, Pat ... I thank Emily Smith and George Hawkswell for introducing me to rowing.

I thank the *Gonville Loungers* for all the great Thursday lounges and the legendary *sans pantalons meales of immense pleasure*: David Gawith, Evan Kaiser, Alex Ross, Nis Kneussels, Jos Crush, Nick Gleadall, Marion ...

I thank all my teachers and supervisors that have taught and guided me throughout the last decade. A special thank goes to Brian Møller Andersen for the years in Copenhagen.

Finally I thank my lovely family. I thank my father for having been a solid reference throughout the times. For shedding light when the road was dark and for pushing just the right amount towards the right direction. I thank my lovely mother for providing me with love. I thank Uma for being the best sister in the world! I thank my *amamas* for being sweet.

Abstract

The interest in vortices and vortex lattices was sparked by the prediction of quantisation of circulation by Onsager in the 1940s. The field has since developed dramatically and attracted a lot of interest across the physics community. In this dissertation we study vortices in two different systems: a rotating, Rabi-coupled, two-component Bose–Einstein condensate (BEC) and a rotating spinor-BEC, in two spatial dimensions.

Vortex molecules can form in a two-component superfluid when a Rabi field drives transitions between the two components. We study the ground state of an infinite system of vortex molecules in two dimensions, using a numerical scheme which makes no use of the lowest Landau level approximation. We find the ground state lattice geometry for different values of intercomponent interactions and strength of the Rabi field. In the limit of large field, when molecules are tightly bound, we develop a complementary analytical description. The energy governing the alignment of molecules on a triangular lattice is found to correspond to that of an infinite system of two-dimensional quadrupoles, which may be written in terms of an elliptic function $\mathcal{Q}(z_{ij};\omega_1,\omega_2)$. This allows for a numerical evaluation of the energy enabling us to find the ground state configuration of the molecules.

In the *polar* phase of a two-component BEC, in which the spin density is zero, the emergent *spin-gauge rotation symmetry* of the order parameter allows for the presence of half-quantum vortices (HQVs). We numerically search for this object in the variational ground state of a spinor-BEC and find it in certain region of the phase diagram. We provide analytical arguments that suggest that this object is energetically favorable in the ground state.

Matrix product state (MPS) based methods are currently regarded as one of the most powerful tools to study the low-energy physics of one-dimensional many-body quantum systems. In this work we find a connection between MPS in the left canonical form and the Stiefel manifold. This relation allows us to constrain the optimisation to this subspace of the otherwise larger MPS manifold. We find that our method suffers from two undesirable features. First, the need of a large unit cell to achieve machine precision. Second, because of the presence of the power method in the variational energy expression, it is possible for the convergence process to get stuck in regions of the Stiefel manifold where the modulus of the second largest eigenvalue of the transfer matrix is very close to one.

Since the foundation of the field of *Artificial Intelligence* (AI) in 1956, at a workshop held in Dartmouth College (New Hampshire, US), the excitement and optimism towards it

has oscillated throughout the years. The last AI boom started in 2012 and we live in a time where people from *all* disciplines, both in industry and academia, are getting involved in machine learning. We contribute to the field with a *quantum-inspired* generative model for raw audio. Our model is based on continuous matrix product states and it takes the form of a *stochastic Schrödinger equation*, describing the continuous time measurement of a quantum system. We test our model on three different synthetic datasets and we find its performance promising.

Contents

1 Introduction

In this chapter we introduce the concepts necessary for the comprehension of the research presented in the subsequent chapters, alongside some historical background.

1.1 Vortices in superfluids

1.1.1 Historical introduction

Quantized vortices have long been understood to be a characteristic of superfluid flow. As stated by Russel J. Donnelly in Ref. [1], the idea of quantized circulation was first announced to students and colleagues at Yale University by Lars Onsager in 1946, in the context of superfluid helium. Onsager said: 'Thus the well-known invariant called the hydrodynamic circulation is quantized; the quantum of circulation is h/m ... In the case of cylindrical symmetry, the angular momentum per particle is a multiple of $\hbar$'[2].

This story was followed by London's work on flux quantization in superconductors [3]. Building on these ideas, in 1955 Feynman came up with the concept of a vortex filament (or line vortex) and suggested that the ground state of a rotating superfluid should be a lattice of vortex filaments. The circulation around each of the vortex lines would be quantized, as Onsager had previously conjectured [4]. Hesitant about the suggested vortex lattice ground state, Feynman says: 'It is not self-evident that there is no state of appreciably lower energy'. Two years later, Abrikosov gave a quantitative theory of the vortex lattice in Type-II superconductors [5][†]. The experimental confirmation of these predictions arrived shortly afterwards [6, 7].

The discussion of vortices in superfluids naturally carries over to the context of Bose-Einstein condensates. Soon after the observation of superfluidity in liquid ^{4}He in 1938 by the group of Kapitza in Moscow and Allen and Misener in Cambridge, London suggested that the phenomena could be related to Bose-Einstein condensation. Superfluid liquid ^{4}He is

[†]In his Nobel Prize lecture, Abrikosov states that his discovery dates from 1953 but publication was delayed by an incredulous Landau!

the prototype BEC, and it has played a unique role in the development of physical concepts. However, the fact that interactions in liquid helium dramatically reduce the occupancy of the lowest single-particle state, led to the search for weakly interacting Bose gases with higher condensate fraction. Alkali atoms were used in the first successful experiments to produce a gaseous BEC. The original experiment was done on a gas of rubidium atoms [8]. Subsequently, Bose-Einstein condensation has been realized experimentally in dilute gases of hydrogen, lithium, sodium, potassium, chromium and cesium, among others.

Following the discovery of Bose-Einstein condensation in atomic gases, much work has been devoted to the properties of rotating condensates in traps; these developments have been reviewed in [9, 10]. Furthermore, there has been extensive interest in the study of vortices in BECs [11].

1.1.2 Vortices in scalar BEC

In the first part of this section we follow Chapter 6 on *Theory of the condensed state* in [12]. In the presence of weak interactions, mean field theory is a valid approach to study the ground state of a Bose gas. The Gross-Pitaevskii equation describes the zero temperature properties of the non-uniform Bose gas when the scattering length a is much less than the mean interparticle spacing.

Within this framework, two bosons interact via a repulsive contact interaction $U\delta(\mathbf{r}-\mathbf{r}')$, where $U = \frac{4\pi\hbar^2 a}{m}$ is the effective interaction at low energies in momentum space. The many-body wavefunction is assumed to be a symmetrized product of single-particle wavefunctions. In the fully condensed state, all bosons are in the same single particle state $\phi(\mathbf{r})$ and the N particle wavefunction is

$$\Psi(\mathbf{r}_1, ..., \mathbf{r}_N) = \prod_{i=1}^{N} \phi(\mathbf{r}_i). \tag{1.1}$$

The relevant low energy Hamiltonian is $H = H_0 + H_{int}$ where

$$\begin{aligned} H_0 &= \int d\mathbf{r}\, \Psi^\dagger(\mathbf{r}) \left[\frac{\mathbf{p}^2}{2m} + V(\mathbf{r}) \right] \Psi(\mathbf{r}), \\ H_{\text{int}} &= \frac{U}{2} \int d\mathbf{r}\, \Psi^\dagger(\mathbf{r}) \Psi^\dagger(\mathbf{r}) \Psi(\mathbf{r}) \Psi(\mathbf{r}). \end{aligned} \tag{1.2a}$$

Defining the condensate wavefunction $\psi(\mathbf{r}) = \sqrt{N}\phi(\mathbf{r})$ and neglecting terms of order $1/N$,

one arrives to the Gross-Pitaevskii energy functional

$$E(\psi) = \int d\mathbf{r} \left(\frac{\hbar^2}{2m} |\nabla\psi(\mathbf{r})|^2 + V(\mathbf{r})|\psi(\mathbf{r})|^2 + \frac{U}{2}|\psi(\mathbf{r})|^4 \right). \tag{1.3}$$

Constraining the number of particles in the condensate to be N, i.e.,

$$N = \int d\mathbf{r}\, |\psi(\mathbf{r})|^2, \tag{1.4}$$

the wavefunction that minimizes the energy is a solution to the *time-independent Gross-Pitaevskii* equation

$$\frac{-\hbar^2}{2m}\nabla^2\psi(\mathbf{r}) + V(\mathbf{r})\psi(\mathbf{r}) + U|\psi(\mathbf{r})|^2\psi(\mathbf{r}) = \mu\psi(\mathbf{r}). \tag{1.5}$$

Here μ is the chemical potential, which is the Lagrange multiplier that imposes the constraint on the number of particles. This equation has the form of a Schrödinger equation, where the potential energy depends on the density of particles, i.e. it takes into account the mean field produced by other the bosons. Note that in this case the eigenvalue is the chemical potential.

While the equilibrium structure of the condensate is described by the *time-independent Gross-Pitaevskii* equation, the dynamics can be described by its time-dependent generalization: the *time-dependent Gross-Pitaevskii* equation

$$i\hbar\frac{\partial\psi(\mathbf{r},t)}{\partial t} = \frac{-\hbar^2}{2m}\nabla^2\psi(\mathbf{r},t) + V(\mathbf{r})\psi(\mathbf{r},t) + U|\psi(\mathbf{r},t)|^2\psi(\mathbf{r},t). \tag{1.6}$$

From this equation, a continuity equation for the condensate is derived

$$\frac{\partial n}{\partial t} + \nabla\cdot(n\mathbf{v}) = 0, \tag{1.7}$$

where $n = |\psi|^2$ and the velocity of the condensate is

$$\mathbf{v} = \frac{\hbar}{2mi}\frac{(\psi^*\nabla\psi - \psi\nabla\psi^*)}{|\psi|^2}. \tag{1.8}$$

Parameterizing the wavefunction as $\psi = fe^{i\phi}$, the velocity is

$$\mathbf{v} = \frac{\hbar}{m}\nabla\phi. \tag{1.9}$$

One of the hallmarks of a superfluid is its response to rotation or, in the case of charged superfluids, the response to a magnetic field. The special properties of superfluids are a consequence of their motion being constrained by the fact that the velocity is proportional to the gradient of the phase of the wavefunction as seen in Eq. (1.9). This form of the

velocity implies that the flow is irrotational, i.e. $\nabla \times \mathbf{v} = 0$. In a simply connected region this has only one solution for a given boundary condition. For $\mathbf{v}_{boundary} = 0$, the solution is $\mathbf{v} = 0$. In a multiply connected region the situation is different. Since $\nabla \times \mathbf{v} = 0$, the circulation about any closed curve that can be shrunk to a point is zero. On the other hand, if the curve encloses a hole, the circulation does not need to vanish. Since the wavefunction needs to be single valued, ϕ can only increase by multiples of 2π going around the hole and so one arrives to the *Onsager-Feynman quantization condition*

$$\oint \nabla\phi \cdot d\mathbf{s} = 2\pi n. \tag{1.10}$$

The presence of a BEC implies the existence of a condensate wavefunction, which requires that the increment of the phase of the wavefunction over a closed path must be an integral multiple of 2π. A vortex is a configuration where the increment is non zero.

In a trap with axial symmetry, a centered vortex wavefunction takes the form (in polar coordinates)

$$\psi(r,\theta) = e^{in\theta} f(r), \tag{1.11}$$

where n is the winding number. In the absence of a trap potential and setting $m = \hbar = 1$, the Gross-Pitaevskii energy of this state is

$$E = \int d\mathbf{r} \left[\frac{1}{2} \left(\frac{\partial f(r)}{\partial r} \right)^2 + \left(\frac{n^2}{2r^2} - \mu \right) f^2(r) + \frac{U}{2} f^4(r) \right] \tag{1.12}$$

and the corresponding Euler-Lagrange equation

$$\frac{-1}{2r} \frac{\partial}{\partial r} \left(r \frac{\partial f(r)}{\partial r} \right) + \left(\frac{n^2}{2r^2} - \mu \right) f(r) + U f(r)^3 = 0. \tag{1.13}$$

The solution to this equation has a characteristic length $\xi = \mu^{-1/2}$ called *healing length*. The amplitude of the wavefunction reaches its bulk value $f(r \to \infty) = \sqrt{\frac{\mu}{U}}$ over a distance ξ from the vortex center.

In the presence of many vortices, the kinetic energy gives rise to a term that corresponds to a system of point charges interacting via 2D Coulomb interactions (see Appendix A.6)

$$E = -2\pi \sum_{i<j} n_i n_j \log \left(\frac{|\mathbf{r}_i - \mathbf{r}_j|}{\xi} \right). \tag{1.14}$$

This term governs the structure of vortex lattices in scalar Bose-Einstein condensates.

1.1.3 Vortices in spinor BEC

In this section we follow Chapter 12 on *Mixtures and spinor condensates* in [12]. Bose-Einstein condensation can also take place by macroscopically occupying two or more quantum states. The simplest example of such a multi-component system is a mixture of two different species of bosons, for example two isotopes of the same element or two different atoms.

The fact that alkali atoms have spin, gives the possibility to make mixtures of the same isotope, but in different internal spin states. This was first done experimentally by the JILA group, where a mixture of ^{87}Rb atoms in the hyperfine states $F = 2, m_F = 2$ and $F = 1, m_F = -1$ was made [13]. Here $\mathbf{F} = \mathbf{I} + \mathbf{J}$, where $\mathbf{I}$ is the nuclear spin and $\mathbf{J}$ is the electronic total angular momentum. Mixtures of hyperfine states of the same isotope are different from mixtures of distinct isotopes in that there is inter-hyperfine state scattering for the first, while transitions that convert one isotope into another may be neglected. When the interaction energy per particle is comparable with or larger than the energy difference between hyperfine levels, transitions between different hyperfine states can influence equilibrium properties. In magnetic traps, because the trapping potential depends on the particular hyperfine state, it is difficult to achieve such conditions. On the other hand, in optical traps it is possible to have magnetic degrees of freedom because the potential is independent of the hyperfine state. The dynamics of the spin has been investigated in these kind of systems, beginning with the experiments of [14]. To study a multi-component BEC, one can generalize the theory used to investigate scalar condensates to allow for the spinor nature of the condensate wavefunction.

The two-component generalization of the Hartree wavefunction shown in Eq. (1.1) is

$$\Psi(\mathbf{r}_1, ..., \mathbf{r}_{N_1}; \mathbf{r}'_1, ..., \mathbf{r}'_{N_2}) = \prod_{i=1}^{N_1} \phi(\mathbf{r}_i) \prod_{j=1}^{N_2} \phi(\mathbf{r}'_j), \tag{1.15}$$

where the particles of species 1 are denoted by $\mathbf{r}_i$ and those of species 2 by $\mathbf{r}'_j$. The relevant low energy Hamiltonian is $H = H_0 + H_{int}$, where

$$\begin{aligned} H_0 &= \sum_{\sigma} \int d\mathbf{r}\, \Psi^\dagger_\sigma(\mathbf{r}) \left[\frac{\mathbf{p}^2}{2m_\sigma} + V_\sigma(\mathbf{r}) \right] \Psi_\sigma(\mathbf{r}), \\ H_{\text{int}} &= \sum_{\sigma_1\sigma_2} \frac{g_{\sigma_1\sigma_2}}{2} \int d\mathbf{r}\, \Psi^\dagger_{\sigma_1}(\mathbf{r}) \Psi^\dagger_{\sigma_2}(\mathbf{r}) \Psi_{\sigma_2}(\mathbf{r}) \Psi_{\sigma_1}(\mathbf{r}). \end{aligned} \tag{1.16a}$$

If we introduce the condensate wavefunctions for the two components according to the

definitions $\psi_a = \sqrt{N_a}\phi_a$ and $\psi_b = \sqrt{N_b}\phi_b$, the Gross-Pitaevskii energy of the mixture is

$$E(\psi_a, \psi_b) = \sum_\sigma \int d\mathbf{r}\, \psi^*_\sigma(\mathbf{r}) \left[\frac{\mathbf{p}^2}{2m_\sigma} + V_\sigma(\mathbf{r}) \right] \psi_\sigma(\mathbf{r}) + \sum_{\sigma_1\sigma_2} \frac{g_{\sigma_1\sigma_2}}{2} \int d\mathbf{r}\, |\psi_{\sigma_1}(\mathbf{r})|^2 |\psi_{\sigma_2}(\mathbf{r})|^2, \tag{1.17}$$

where we have neglected terms of order $1/N_a$ and $1/N_b$, which are small when N_a and N_b are large. Here m_σ is the mass of an atom of species σ, and V_σ is the external potential. The constants g_{aa}, g_{bb} and $g_{ab} = g_{ba}$ are related to the respective scattering lengths a_{aa}, a_{bb} and $a_{ab} = a_{ba}$, by $g_{ij} = 2\pi\hbar^2 a_{ij}/m_{ij}$, where $m_{ij} = m_i m_j/(m_i + m_j)$ is the reduced mass for an atom i and an atom j.

In the mixtures considered above, the number of particles per component is conserved, the interactions do not mediate inter-component transitions. The situation is different when the condensation occurs in different hyperfine states of same isotopes. The initial experiments on overlapping condensates with atoms in two different hyperfine states were made in magnetic traps [13]. It was the development of purely optical traps that allowed to condense Na atoms in a superposition of three magnetic substates, corresponding to the quantum numbers $m_F = 0, \pm 1$, of the hyperfine multiplet with total spin $F = 1$ [14].

Let us consider magnetic fields so low that there is rotational invariance such that the states of a single atom are eigenstates of the angular momentum. Rotational invariance imposes important constraints on the number of parameters needed to characterise the interaction. Two identical bosonic atoms with $F = 1$ in an s-state ($L = 0$) of the relative motion can couple to make states with total angular momentum $\mathcal{F} = 0$ or 2 ($|F_1 - F_2| \leq \mathcal{F} \leq F_1 + F_2$), since the possibility of unit angular momentum is ruled out by the requirement that the wavefunction be symmetric under exchange of the two atoms (s-wave and boson $\Rightarrow$ symmetric spin state). Using this knowledge, one can simplify the interaction term. The interaction is invariant under rotations, and therefore it is diagonal in the total angular momentum of the two atoms. One can then write the effective interaction for low-energy collisions as $U_0 = 4\pi\hbar^2 a_0/m$ for $\mathcal{F} = 0$ and $U_2 = 4\pi\hbar^2 a_2/m$ for $\mathcal{F} = 2$, where a_0 and a_2 are the corresponding scattering lengths.

For the purpose of calculating low-energy properties of the system, the true interatomic potential, which in general has a complicated dependence on the inter-particle separation, may be replaced by an effective interaction that is proportional to the scattering length. In real space, the effective interaction between two particles may be taken to be contactlike

$$U_{eff}(\mathbf{r}_1, \mathbf{r}_2) = U_\mathcal{F} \delta(\mathbf{r}_1 - \mathbf{r}_2). \tag{1.18}$$

For two atoms with $F = 1$, the effective interaction may therefore be written in the form

$$\hat{U}_{\mathcal{F}} = \delta(\mathbf{r}_1 - \mathbf{r}_2)(U_0\mathcal{P}_0 + U_2\mathcal{P}_2),$$

where $\mathcal{P}_\sigma = |\sigma\rangle\langle\sigma|$. We can rewrite this result in terms of the operators for the angular momentum of the two atoms $\mathbf{F}_1$ and $\mathbf{F}_2$. Firstly, note that

$$\mathcal{P}_0|\mathcal{F}\rangle = \delta_{0,\mathcal{F}}|0\rangle = \delta_{0,\mathcal{F}}|\mathcal{F}\rangle \text{ and } \mathcal{P}_2|\mathcal{F}\rangle = \delta_{2,\mathcal{F}}|2\rangle = \delta_{2,\mathcal{F}}|\mathcal{F}\rangle. \tag{1.19}$$

Furthermore,

$$\mathbf{F}_1 \cdot \mathbf{F}_2|\mathcal{F}\rangle = \frac{1}{2}(\mathcal{F}^2 - \mathbf{F}_1^2 - \mathbf{F}_2^2)|\mathcal{F}\rangle = \frac{\hbar^2}{2}(\mathcal{F}(\mathcal{F}+1) - 4)|\mathcal{F}\rangle. \tag{1.20}$$

By inspection, one can see that since $\mathcal{F} = 0$ or 2,

$$\frac{(1 - \mathbf{F}_1 \cdot \mathbf{F}_2/\hbar^2)}{3}|\mathcal{F}\rangle = \delta_{0,\mathcal{F}}|\mathcal{F}\rangle \text{ and } \frac{(2 + \mathbf{F_1} \cdot \mathbf{F}_2/\hbar^2)}{3}|\mathcal{F}\rangle = \delta_{2,\mathcal{F}}|\mathcal{F}\rangle. \tag{1.21}$$

Hence, we identify

$$\mathcal{P}_0 = \frac{(1 - \mathbf{F}_1 \cdot \mathbf{F}_2/\hbar^2)}{3} \text{ and } \mathcal{P}_2 = \frac{(2 + \mathbf{F}_1 \cdot \mathbf{F}_2/\hbar^2)}{3}. \tag{1.22}$$

The effective interaction is then

$$\hat{U}_{\mathcal{F}} = \delta(\mathbf{r}_1 - \mathbf{r}_2)(W_0 + W_2\mathbf{F}_1 \cdot \mathbf{F}_2),$$

where

$$W_0 = \frac{U_0 + 2U_2}{3} \text{ and } W_2 = \frac{U_2 - U_0}{3\hbar^2}. \tag{1.23}$$

Due to the off-diagonal coupling stemming from the spin-spin interactions, a suitable ground state *ansatz* is a condensate in a superposition of the different hyperfine states

$$|\Psi\rangle = \frac{1}{(N!)^{1/2}}\left(\sum_\sigma \int d\mathbf{r}\, \phi_\sigma^*(\mathbf{r})\Psi_\sigma^\dagger(\mathbf{r})\right)^N |0\rangle. \tag{1.24}$$

The Gross-Pitaevskii energy of this state is

$$E(\Psi) = \int d\mathbf{r}\left(\frac{\hbar^2}{2m}\partial_i\Psi^\dagger \cdot \partial_i\Psi - \mu\Psi^\dagger \cdot \Psi + \frac{W_0}{2}(\Psi^\dagger \cdot \Psi)^2 + \frac{W_2}{2}(\Psi^\dagger\mathbf{F}\Psi)^2\right), \tag{1.25}$$

where we have defined the spinor $\Psi \equiv (\psi_1, \psi_0, \psi_{-1})^T$.

1.2 Matrix product states

This section follows closely the introduction to matrix product states by Román Orus in Ref. [15]. One of the most challenging problems in condensed matter physics is the study of interacting quantum systems. The clearest manifestation of this difficulty is the yet undiscovered mechanism which gives rise to high-temperature superconductivity, which has eluded physicists since its discovery in 1986 by IBM researchers Georg Bednorz and K. Alex Müller, who were awarded the 1987 Nobel Prize in Physics "for their important break-through in the discovery of superconductivity in ceramic materials".

The standard approach when studying these systems is to construct a simplified model that is believed to capture the relevant correlations, e.g., the Hubbard model for high-T_c cuprates. In most cases, one needs to rely on faithful numerical methods to determine their properties. In some lucky cases, these models are exactly solvable, although these solutions are often limited to systems in one spatial dimension.

Tensor networks have become very popular in recent years as a tool for numerical simulation of quantum many-body systems [16]. The main idea behind this method is to decompose the wavefunction, considered as a tensor in an exponentially large Hilbert space, into smaller tensors that are interconnected. The required size of the tensors is directly connected to the amount of entanglement present in the many-body state. One-dimensional tensor networks are called *matrix product states* (MPS).

1.2.1 The Hilbert space is humongous!

The Hilbert space of a many-body system is exponentially large in the number particles. For example, if we consider N spin-$1/2$ particles, the dimension of the Hilbert space is 2^N. If we consider a macroscopic object where the number of constituents is $\sim 10^{23}$ (of the order of the Avogadro number) the number of basis states in the Hilbert space is $\sim O(10^{10^{23}})$, which is exponentially larger than the number of atoms in the observable universe, estimated to be around 10^{80}!

Luckily for theoretical physicists, not all the states in the Hilbert space are equally relevant, if we are interested in understanding low energy properties of the system. Concretely, many Hamiltonians that describe real systems tend to have short-range interactions, i.e., local interactions. The locality of interactions has important implications, namely, one can prove that low energy states of local Hamiltonians with a gap between the ground state and the first excited state, obey the so called *area law for the entanglement entropy*

[17]. This means that the entanglement entropy between some region of the system and its complement tends to scale, for sufficiently large regions, as the area L^{d-1} of the boundary of the region and not as the volume enclosed L^d. Therefore low-energy states of local, gapped Hamiltonians live in a small manifold of the gigantic Hilbert space, the manifold where the area law is fulfilled. The family of tensor network states are the natural parameterization of this manifold. These states are atypical because a quantum state picked at random from a many-body Hilbert space will most likely have a volume law. The manifold containing area-law states is an exponentially small corner of the Hilbert space.

1.2.2 Breaking the wavefunction into small pieces

Let us consider a quantum many-body system of N particles in one dimension. Each particle has d degrees of freedom. In general, a state of this system is expressed as

$$|\Psi\rangle = \sum_{s_1,...,s_N=1}^{d} C_{s_1,...,s_N} |s_1, ..., s_N\rangle , \tag{1.26}$$

where $\{|s_1, ..., s_N\rangle\}$ is a basis of choice. The expansion coefficients $\{C_{s_1,...,s_N}\}$ can be seen as being elements of a tensor of rank N. This tensor contains, in general, d^N independent, complex coefficients (up to normalization), i.e., an exponentially large number of coefficients. One of the aims of using MPSs is to reduce the complexity of the representation of states like $|\Psi\rangle$ whilst still providing an accurate description of the entanglement properties of the state. This is achieved by replacing the tensor C by a network of tensors $\{A^s\}$ that have a smaller rank, so that

$$|\Psi_{MPS}\rangle = \sum_{s_1,...,s_N=1}^{d} A^{s_1}...A^{s_N} |s_1, ..., s_N\rangle . \tag{1.27}$$

Note that by making the matrices A^s large enough, one recovers the complexity of the tensor C. Indeed, any 1D state can be represented exactly using an MPS [18]. However, it turns out that the states that obey the area law can be successfully represented by A^s that are small enough to make $A^{s_1}...A^{s_N}$ contain a lot fewer coefficients than C. As it turns out, $|\Psi_{MPS}\rangle$ typically depends on a polynomial number of parameters, therefore being a computationally feasible description of a many-body state.

The indices that connect the matrices in the network have an important physical meaning: they represent the structure of the many-body entanglement in the state. The number of values that each of one of these indices can take is directly related to the amount of entanglement in the wavefunction. These indices are called *bond indices*. The maximum of these values is called the *bond dimension D*. In the case of a translationally invariant

MPS, all A^s matrices are equal and $|\Psi_{MPS}\rangle$ is represented by dD^2 complex numbers, i.e., independent of the system size.

1.2.3 Continuous matrix product states

The physical lattice states are well-captured by matrix product states. In the context of continuous quantum systems, there exist a continuum limit, without any reference to an underlying lattice parameter. This family of states are called *continuous matrix product states* (cMPS). They describe field theories in one spatial dimension. Just as MPS captures the entanglement structure of low energy states of quantum spin systems, the entanglement structure of cMPS is tailored to describe the low-energy states of quantum field theories [19].

In order to define the cMPS, let us consider a one-dimensional system of bosons or fermions on a ring of length L and associated field operators $\hat{\psi}(x)$ with canonical commutation relations, $[\hat{\psi}(x), \hat{\psi}(y)^\dagger] = \delta(x-y)$ with $0 \leq x, y \leq L$ space coordinates. A cMPS is defined as

$$|\Psi\rangle = \mathrm{Tr}_{aux}\left[\mathcal{P}e^{\int_0^L dx\left[Q(x)\otimes\mathbb{1}+R(x)\otimes\hat{\psi}(x)^\dagger\right]}\right]|0\rangle\,, \tag{1.28}$$

where $Q(x)$, $R(x)$ are position dependent matrices of dimension $D \times D$ that act on a D-dimensional auxiliary system, $\mathcal{P}$ exp is the notation for the path-ordered exponential, Tr_{aux} is the trace over the auxiliary system and $|0\rangle$ is the vacuum state $\hat{\psi}(x)|0\rangle = 0$. A translationally invariant state corresponds to $Q(x)$ and $R(x)$ being independent of x, and a system with open boundary conditions can be obtained by replacing the Tr_{aux} by a left and right multiplication of the auxiliary system with a row and a column vector, respectively:

$$|\Psi\rangle = \langle v_L|\,\mathcal{P}e^{\int_0^L dx\left[Q(x)\otimes\mathbb{1}+R(x)\otimes\hat{\psi}(x)^\dagger\right]}\,|v_R\rangle \otimes |0\rangle\,. \tag{1.29}$$

1.2.4 Connection between MPS and cMPS

As shown by Schön *et al.* in Ref. [20], an MPS with bond dimension D can be seen as a sequentially generated multiqubit state, arising from a D-level system. Let $\mathcal{H}_A = \mathbb{C}^D$ and $\mathcal{H}_B = \mathbb{C}^2$ be the Hilbert spaces of the ancilla and a single qubit respectively. In every step of the sequential generation, we consider unitary evolution of the joint system $\mathcal{H}_A \otimes \mathcal{H}_B$. Assuming that each qubit is initially empty $|0\rangle$, we disregard the qubit at the input, such that the evolution takes the form of an isometry $V:\ \mathcal{H}_A \to \mathcal{H}_A \otimes \mathcal{H}_B$. Choosing a basis in

the ancilla space, the isometry is expressed as

$$V = \sum_s \sum_{a,b} A^s_{a,b} \left(|a\rangle \langle b| \otimes |s\rangle \right), \tag{1.30}$$

where $\sum_s A^{s\dagger} A^s = \mathbb{1}$ is the isometry condition and each A^s is a $D \times D$ matrix. After applying V n times to an initial state $|\psi_\mathrm{I}\rangle \in \mathcal{H}_A$,

$$|\Psi\rangle = \sum_{\mathbf{s}} \sum_{ab} A^{s_n}_{a,} ... A^{s_1}_{,b} \langle b|\psi_\mathrm{I}\rangle \left(|a\rangle \otimes |\mathbf{s}\rangle \right). \tag{1.31}$$

The generated n qubits are in general entangled both with the ancilla and between themselves. If the ancilla is decoupled in the last step, the final state is an MPS in the space of the n qubits:

$$|\Psi\rangle = |\psi_\mathrm{F}\rangle \otimes \sum_{\mathbf{s}} \sum_{ab} \langle \psi_\mathrm{F}|a\rangle A^{s_n}_{a,} ... A^{s_1}_{,b} \langle b|\psi_\mathrm{I}\rangle |\mathbf{s}\rangle . \tag{1.32}$$

This result shows that all sequentially generated multiqubit states, arising from a D-dimensional ancillary system $\mathcal{H}_A$, are instances of MPS with $D \times D$ matrices A^s and open boundary conditions specified by $|\psi_\mathrm{I}\rangle$ and $|\psi_\mathrm{F}\rangle$.

Let us now consider the cMPS shown in Eq. (1.29), without projecting the ancilla onto $|v_L\rangle$. Taking $L = dx$ and Q, R translationally invariant,

$$\begin{aligned}
|\Psi\rangle &= \mathcal{P} e^{dx\left[Q\otimes\mathbb{1} + R\otimes\hat{\psi}(x)^\dagger\right]} |v_R\rangle \otimes |0\rangle \\
&= \left[\mathbb{1}\otimes\mathbb{1} + Q dx \otimes \mathbb{1} + R dx \otimes \hat{\psi}(x)^\dagger\right] |v_R\rangle \otimes |0\rangle \\
&= \sum_{ab} \left[(\delta_{ab} + Q_{ab} dx) |a\rangle |b\rangle \otimes \mathbb{1} + R_{ab} dx |a\rangle |b\rangle \otimes \hat{\psi}(x)^\dagger \right] |v_R\rangle \otimes |0\rangle \\
&= \sum_s \sum_{ab} A^s_{ab} \left(|a\rangle \langle b| \otimes |s\rangle \right) |v_R\rangle ,
\end{aligned} \tag{1.33}$$

where $A^0_{ab} = \delta_{ab} + Q_{ab} dx$, $A^1_{ab} = R_{ab} dx$ $\psi^{\dagger 0}(x) = \mathbb{1}$ and $\psi^{\dagger s}(x)|0\rangle = |s\rangle$. This is just the isometry shown in Eq. (1.30).

1.3 Machine learning

In this section we are going to go through a basic introduction to machine learning. The parts that are most relevant to the research shown in this thesis are the sections 1.3.1.1, 1.3.3, 1.3.4, 1.3.4.4,1.3.6 and 1.3.7.3.

There has been a long-standing interest in intelligent machines, nowadays more commonly known as *Artificial Intelligence*. With the advent of the *Third Industrial revolution* (also called the *Digital Revolution*), what before belonged to the realm of science-fiction has been now turned into one of humanity's greatest endeavors. Science and technology are instrumental in making human life better, either contributing to our survival (e.g. medicine) or bringing joy (e.g. smartphones). Both interests are legitimate and serve as an engine for us humans to continue pushing the boundaries of our knowledge.

The field of *Deep learning* and the use of *neural networks* for machine learning, is partly motivated by the fact that humans are extremely good at learning. The infant brain is capable of solving problems that even the most powerful computers find impossible. Soon after the birth, infants are able to recognize their parents faces and discern discrete object from their background. Within a few months, they have developed an intuition for natural physics, can track objects even when they become partially or completely blocked, and can associate sounds with specific meanings [21]. Hence, the desired machine is one that has the human ability to learn with the computing capabilities of a computer.

The fundamental difference between traditional computer programs and the machine learning approach is that the first consist on following a set of instructions to solve the task at hand, whereas in the second, there is also a set of instructions (after all humans need to do the programming) but one can say that these are instructions on how to learn to solve the task, rather than instructions on how to solve it. This different philosophy of solving problems might seem awkward, but the following two examples illustrate well why might it be a better approach for certain tasks.

Imagine the task of having a computer recognize hand written digits. One could try and implement the rules to discern between different digits, but this would be incredibly hard and the amount of rules that one would need to implement would be massive. Instead, the machine learning approach would be to parameterize a function that maps the space of pixels to the space of digits. We call this function the *model*. Furthermore, we would design an objective function called the *loss* function, that measures the correctness of the mapping. Then the task would be to minimise this function to optimise the performance of the model. The process of optimisation is called the *training*.

Another example is the task of making a Bach music generator. A machine that would be able to create new music, that would be recognized as being Bach music by an expert that was unaware of the fact that he died in 1750. It would be rather complicated to come up with a list of rules that would enclose the features that characterize Bach music. In the machine learning approach, one would parametrize the space of different songs and design a loss function that would measure how close this space is from the space of Bach songs. Again the task would be to optimise this function.

Designing the *model* and the *loss* function is an art. The loss function involves both the *parameters* to be learned and the *dataset* used to help choose the optimal parameters. Depending on whether the dataset contains *labels* or not, we call the optimization process *supervised learning* or *unsupervised learning*, respectively. The hand written digit recognition task belongs to the first kind. The dataset contains the hand written digits as well as labels that specify the digits. The Bach music generator task belongs to the second kind, where there are no labels.

In *Deep learning*, part of the inspiration to build the model comes from the human brain. A detailed analysis of the structure of *neural networks* can be found in [21].

In this section we have given a qualitative description of how machine learning works. In the next two sections, we will give detailed examples that involve the mathematics used.

1.3.1 Artificial neural networks

The building block of the brain is the neuron. A tiny piece of the size of a grain of rice of the human brain contains around 10,000 neurons. Each neuron is connected to approximately 6,000 other neurons. Each of these, contains several antennae called *dendrites*, from which it receives inputs. The strength of each input is dynamically strengthen or weakened depending on how much it is used. The inputs are summed together inside the neuron and transformed into a new signal that is propagated along the *axons*. Fig. 1.1 illustrates the structure of a biological neuron.

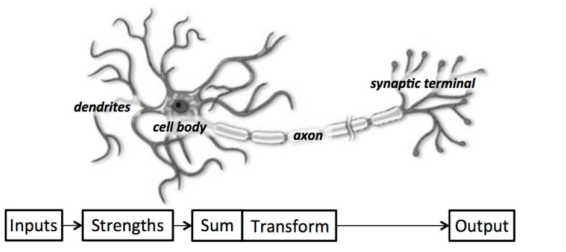


Figure 1.1: Structure of a biological neuron. Figure taken from Ref. [21].

Based on this knowledge, we can make a model in a computer that has the structure of a neuron. Our model neuron will consist of a set of inputs $\{x_i\}$ multiplied by a set of *weights* $\{w_i\}$. See Fig. 1.2.

The weighted inputs are summed together forming the *activation* $a = \sum_i w_i x_i$. Sometimes a constant b called the *bias* is added to the activation. The output of the neuron is a function of the activation $y = f(a)$. It can be shown that non-linear neurons (i.e., f is non-linear) are more powerful than linear neurons [21], therefore they are usually defined as non-linear functions. Commonly used neurons are

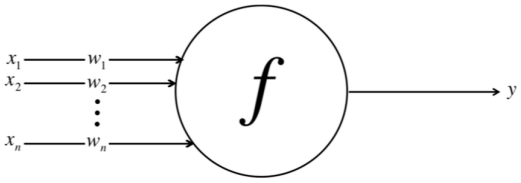


Figure 1.2: Structure of the model neuron. Figure taken from Ref. [21].

$$\text{Sigmoid: } f(a) = \frac{1}{1+e^{-a}}, \tag{1.34}$$

$$\text{Tanh: } f(a) = \tanh(a), \tag{1.35}$$

$$\text{ReLU: } f(a) = \max(0, a). \tag{1.36}$$

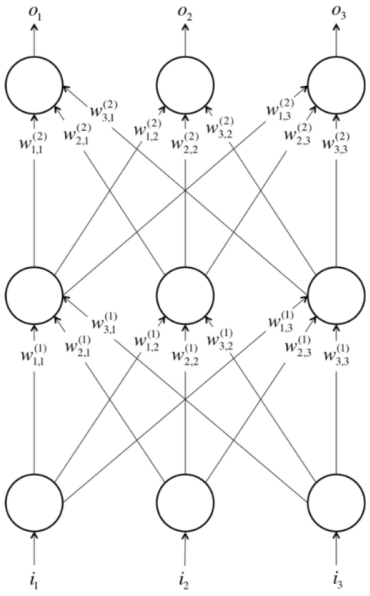


Figure 1.3: A *feed-forward* neural network. Figure taken from Ref. [21].

We construct *artificial neural networks* by connecting different neurons. As an example, see Fig. 1.3. This a *feed-forward* neural network (sometimes called a *multilayer perceptron*) characterized by: 1) all neurons of each *layer* are connected across layers to all the neurons of the next layer, 2) within a layer neurons are not interconnected and 3) information only flows in one direction. This neural network has a *width* of three neurons and a *depth* of three layers: an input layer, a *hidden* layer and an output layer. Different arrangements of connections between layers correspond to different architectures. Amongst the most popular architectures of modern deep learning are *recurrent neural networks* (RNNs) and *convolutional neural networks* (CNNs). See Ref. [21] for a detailed exposition of the CNN

architecture.

1.3.1.1 Recurrent neural networks

The structure of an RNN is shown in Fig. 1.4. Consider the input sequence $\mathbf{x}^{(t)}$. As seen in Fig. 1.4, the function $\mathbf{h}^{(t)}$ called the *hidden state*, evolves in time and its time evolution takes the form of an externally driven dynamical system

$$\mathbf{h}^{(t)} = f(\mathbf{h}^{(t-1)}, \mathbf{x}^{(t)}; \boldsymbol{\theta}), \tag{1.37}$$

where $\boldsymbol{\theta}$ represents the training variables of the neural network, in this case the matrices $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$. The hidden state place the roll of a memory in that it depends on all previous inputs. The sequence $\mathbf{o}^{(t)}$, is the output of the RNN. This architecture is suitable for modeling sequential data. Note that the RNN structure shown in Fig. 1.4 is one concrete structure, but there are other types of RNN. For instance, a possible variation of this model would be an architecture where the output $\mathbf{o}^{(t)}$ was connected to the hidden state $\mathbf{h}^{(t+1)}$.

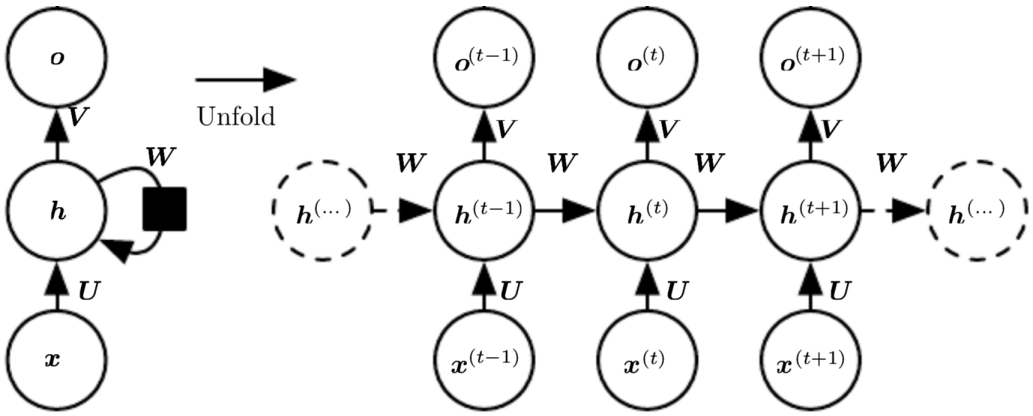


Figure 1.4: Structure of an RNN at training time. The sequence $\mathbf{x}^{(t)}$ is the input data, $\mathbf{h}^{(t)}$ is the hidden state and $\mathbf{o}^{(t)}$ is the output sequence. Figure taken from [22].

1.3.2 The mechanics of *supervised* machine learning

Consider the task where we want to determine how to predict whether a football team will win a game based on the number of hours of training they had and the amount of hours they slept in the last week. We collect a *data set* $\{\mathbf{x}_n\}$ with binary *labels* $\{t_n\}$. Each data point $\mathbf{x} = (x_1, x_2)$ contains the number of training hours (x_1), the number of sleeping hours (x_2) and the outcome of the game: win or lose (1 or 0, respectively).

1.3.2.1 The model

The task is to learn a *model* $f(\mathbf{x}, \mathbf{w}, b)$, such that

$$f(\mathbf{x}, \mathbf{w}, b) = \begin{cases} 1, & \text{if } \mathbf{x} \cdot \mathbf{w} + b \geq 0, \\ 0, & \text{if } \mathbf{x} \cdot \mathbf{w} + b < 0. \end{cases} \tag{1.38}$$

Gometrically, the blueprint of our model $f(\mathbf{x}, \mathbf{w}, b)$ describes a linear classifier that divides the coordinate plane into two halves. We want to *learn* the *parameters* $(\mathbf{w}, b)$ such that our model makes the right predictions given an input $\mathbf{x}$. This model is called the *linear perceptron* [23]. Note that our model has the structure of a *neuron*.

The labels of our data are binary (win or lose), therefore the binary function $f(\mathbf{x}, \mathbf{w}, b)$ is well suited to the task. On the other hand, as we will later see, it is convenient to have a model that is differentiable for *training*. A natural continuous analogous of the linear perceptron is the *sigmoidal neuron*

$$y(\mathbf{x}, \mathbf{w}, b) = \frac{1}{1 + \exp(\mathbf{x} \cdot \mathbf{w} + b)}. \tag{1.39}$$

The domain of this function is $y \in (0, 1)$ and the right choice of $(\mathbf{w}, b)$ can make it arbitrarily close to the linear perceptron.

1.3.2.2 The loss function

We need to design a *loss* function whose minimisation will optimize the performance of our model $y(\mathbf{x}, \mathbf{w}, b)$. An appropriate function is

$$E(\mathbf{w}, b) = \sum_n \left[t_n - y(\mathbf{x}_n, \mathbf{w}, b) \right]^2. \tag{1.40}$$

The sum is over the whole dataset $\{\mathbf{x}_n\}$. This is a good loss function because its lower bound is zero and it only attains this value when the model perfectly matches the data, i.e. $y(\mathbf{x}_n, \mathbf{w}, b) = t_n$ for all n. Therefore, the minimization of this loss function will optimize the performance of the model.

1.3.3 The mechanics of *unsupervised* machine learning

Consider the task of building a device that generates bird sounds. We could just go to the forest and record, but after a while we would get bored of listening to the same sounds once and again. Hence we want a device that can create new bird sounds.

1.3.3.1 The model

We need a generative process to generate the sound. One option is to use the same generative process that generates the sound in the forest, in which case we could buy a few birds in the pet shop and let them tweet. If we are allergic to birds, we might prefer to have our laptop produce the sound. One option is to assume that there exists a probability distribution whose samples are bird sounds. Then the task is to find it. For this purpose one can write down a *model* probability function with a set of *parameters* and try to find the parameters that give rise to the desired sound. We could use an RNN to parameterize the probability distribution. Let us call this function $p(x(t)|\mathbf{w})$, where $\mathbf{w}$ is a vector that contains the *learning variables*. The function $p(x(t)|\mathbf{w})$ is the likelihood of the sound $x(t)$, given the parameters $\mathbf{w}$. This is a *probabilistic generative model*.

1.3.3.2 The loss function

We have a *data set* $\{x_n(t)\}$ of four second clips of bird sounds. A common approach is to *learn* the parameters $\mathbf{w}$ that maximize the likelihood $p(\{x_n(t)\}|\mathbf{w})$ of the data, given the parameters. This is the *Maximum Likelihood* method. The philosophy behind this method is to find the set of parameters $\mathbf{w}^*$ that best fit the data, by maximizing $p(\{x_n(t)\}|\mathbf{w})$.

We assume that the dataset is *independent and identically distributed* (i.i.d). Hence, the joint probability of the whole dataset is the product of the probability of each data point. It is computationally convenient to instead maximise the $\log p(\{x_n(t)\}|\mathbf{w})$ because when the probabilities are very small, the product of them (that appears when the data is i.i.d) is a lot smaller than the sum and so trickier to treat numerically. This method is then called *Maximum Log Likelihood*. Then the logarithm of the joint probability of the whole dataset, is the sum of the log probability of each of the data points

$$E(\mathbf{w}) = -\log \prod_n p(\mathbf{x}_n|\mathbf{w}) = -\sum_n \log p(\mathbf{x}_n|\mathbf{w}). \tag{1.41}$$

1.3.4 Training

Training is the process of optimising the loss function with the aim of making our model successful at a given task. The loss function contains the training variables $\mathbf{w}$ and the training dataset $\{\mathbf{x}_i\}$. Optimizing *deep neural networks* is challenging because of the large amount of learning variables $\mathbf{w}$. Furthermore, large dataset are usually needed to train these networks, which increases the expense of the optimization (note that the loss functions in Eqs. (1.41) and (1.40) involve a sum over the whole training set). Another challenging problem that often appears is the presence of local minima in the loss function

$E(\mathbf{w})$. Local minima are not problematic if the corresponding set of weights $\mathbf{w}^*$ gives rise to a good performance of the model. Local minima are problematic when their corresponding value does not give rise to a good performance of the model. These are called *spurious* local minima.

1.3.4.1 Gradient descent

The most rudimentary optimization algorithm is the *gradient descent* algorithm and all modern refinements build on it. In this method, we calculate the gradient of the loss function with respect to each of the weights in the vector $\mathbf{w}$, in each step. Then, we take a step in the opposite direction of the gradient, which gives the steepest direction in the loss function landscape

$$\Delta\mathbf{w} = -\epsilon\nabla_{\mathbf{w}}E. \tag{1.42}$$

The *hyperparameter* ϵ is the learning rate, which sets the size of the step taken. A *hyperparameter* is a parameter that is part of the *model* or the *loss*, or both. Hyperparameters are not learned, we experiment with different values and pick the one that gives the best performance.

1.3.4.2 Backpropagation

Consider the output layer j of the neural network shown in Fig. 1.5. The neurons are *sigmoidal*

$$y = \frac{1}{1+e^{-a}}, \text{ where } a = \mathbf{w}\cdot\mathbf{x}. \tag{1.43}$$

In general the output y_j is a very complicated function, it is the composition of many non-linear functions. The idea behind the *backpropagation* algorithm is to exploit the chain rule of differentiation to calculate gradients of these composed functions in an efficient manner. To get a hint of the details, let us go through an example. Consider the neural network in Fig.1.5.

Let us show in three steps that for a given layer i, it is possible to express the gradient of the outputs on this layers, in terms of the gradients of the layer above $j = i + 1$.

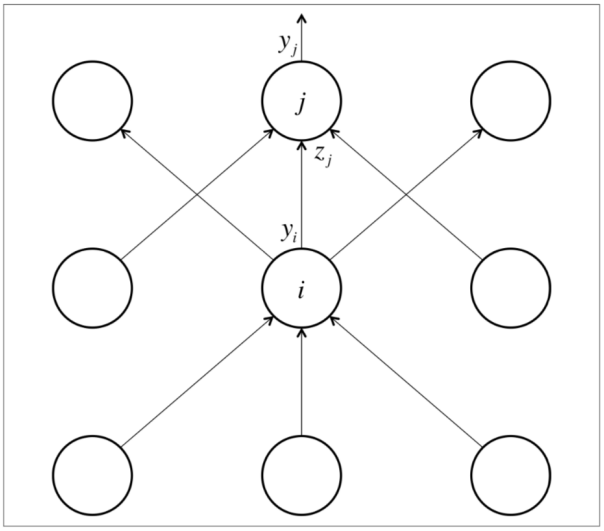


Figure 1.5: A diagram to follow the derivation the backpropagation algorithm. Figure taken from [21].

$$
\begin{aligned}
&1)\ \frac{\partial E}{\partial y_i} = \sum_j \frac{\partial E}{\partial a_j}\frac{\partial a_j}{\partial y_i} = \sum_j w_{ij}\frac{\partial E}{\partial a_j},\\
&2)\ \frac{\partial E}{\partial a_j} = \frac{\partial E}{\partial y_j}\frac{\partial y_j}{\partial a_j} = y_j(1-y_j)\frac{\partial E}{\partial y_j},\\
&3)\ \frac{\partial E}{\partial y_i} = \sum_j w_{ij}y_j(1-y_j)\frac{\partial E}{\partial y_j}.
\end{aligned}
\tag{1.44}
$$

Once we have computed all the gradients with respect to the outputs of each layer, we can compute the gradients with respect to the weights

$$1)\ \frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial y_j}\frac{\partial y_j}{\partial w_{ij}} = \frac{\partial E}{\partial y_j}\frac{\partial y_j}{\partial z_j}\frac{\partial z_j}{\partial w_{ij}}, \tag{1.45}$$

$$2)\ \frac{\partial y_j}{\partial z_j} = y_j(1-y_j), \tag{1.46}$$

$$3)\ \frac{\partial z_j}{\partial w_{ij}} = y_i, \tag{1.47}$$

$$4)\ \frac{\partial E}{\partial w_{ij}} = y_j(1-y_j)y_i\frac{\partial E}{\partial y_j}. \tag{1.48}$$

where $\frac{\partial E}{\partial y_j}$ depends on the specific loss function. Hence we have a method to compute gradients in an efficient manner. Finally, taking into account that the loss function is in general a sum over the loss contribution of each of the data points, we need to some over

all contributions

$$\Delta w_{ij} = -\epsilon \sum_n y_i^{(n)} y_j^{(n)} (1 - y_j^{(n)}) \frac{\partial E^{(n)}}{\partial y_j}. \tag{1.49}$$

If for each update, we consider the whole dataset, this is called *batch gradient descent*. In practice this is never done, mainly because it is computationally unaffordable for large dataset. Secondly, because this method tends to get stuck in *spurious* minima. The usual approach is consider one subset of the dataset at a time. This method is called *mini-batch gradient descent*. The size of the *mini-batch* is a hyperparameter.

Autodifferentiation

Here we follow the explanation given in Ref. [24]. Autodifferentiation (or automatic differentiation) is a set of techniques to numerically evaluate the derivative of a function defined in a computer program. This technique exploits the fact that every computer program, regardless of how complicated it is, consists on executing a sequence of elementary arithmetic operations (addition, substraction, multiplication, division, etc.) and elementary functions (exp, log, sin, cos, etc.). By applying the chain rule of differentiation to this to these operations, derivatives of arbitrary order can be computed automatically and accurately to working precision.

Autodifferentiation *is not* numerical differentiation. Unlike in numerical differentiation, there is no discretization process that leads to round-off errors.

Usually the gradients in the backpropagation algorithm are calculated using autodifferentiation.

1.3.4.3 *Adam*

Throughout the past few years, a several refinements of gradient descent have appeared: *momentum gradient descent*, *Adagrad*, *RMSProp*, *Adam* ... The method called *Adam* is currently the one that is most widely used. The update equation is defined as

$$\mathbf{w}_t = \mathbf{w}_{t-1} - \frac{\epsilon}{\sqrt{\tilde{\mathbf{v}}}} \tilde{\mathbf{m}}, \text{ where} \tag{1.50}$$

$$\mathbf{m}_t = \beta_1 \mathbf{m}_{t-1} + (1 - \beta_1) \nabla_{\mathbf{w}} E, \tag{1.51}$$

$$\tag{1.52}$$

and

$$\mathbf{v}_t = \beta_2 \mathbf{v}_{t-1} + (1-\beta_2)\nabla_{\mathbf{w}} E \odot \nabla_{\mathbf{w}} E, \tag{1.53}$$

$$\tilde{\mathbf{m}} = \frac{\mathbf{m}_t}{1-\beta_1^t}, \tag{1.54}$$

$$\tilde{\mathbf{v}} = \frac{\mathbf{v}_t}{1-\beta_2^t}. \tag{1.55}$$

The index t is the step in the iteration process and $\mathbf{v}_0 = \mathbf{m}_0 = 0$. The symbol $\odot$ means element-wise multiplication. The idea behind this method is to damp the wild oscillations that average to zero over many steps, speeding up the convergence process. This oscillating behavior is common when using gradient descent. For a detailed exposition of the method see Ref. [21].

1.3.4.4 Regularization

A problem that often appears is that the model performs well on the data that it was trained on, but when evaluated on new data, the performance is not good enough. This takes us to the concept of *generalization*. A model is said to generalize well, when its performance remains successful when tested on new data. The problem of poor generalization is called *overfitting*.

There are several techniques to prevent overfitting. One of them is called *regularization*. This method consists on modifying the loss function to prevent the weights $\mathbf{w}$ to become very large during training. There are different kinds of regularization methods. The most common type is the *L2 regularization*. This is introduced by adding a term $\frac{1}{2}\lambda \sum_i w_i^2$ to the loss function, where the sum is over all the weights in $\mathbf{w}$. The intuition behind his regularization method is to penalize peaky weight vectors and preferring diffuse weight vectors [21].

Note that in the context of log maximum likelihood, adding this term is equivalent to maximizing our model likelihood times a Gaussian prior of the weights

$$-\frac{1}{2}\lambda \sum_i w_i^2 = -\log \exp\left(-\frac{1}{2(1/\sqrt{\lambda})^2}\sum_i w_i^2\right). \tag{1.56}$$

Using Bayes theorem

$$p(\mathbf{w}|\mathbf{x}) = \frac{p(\mathbf{x}|\mathbf{w})p(\mathbf{w})}{p(\mathbf{x})}, \tag{1.57}$$

and

$$\log p(\mathbf{w}|\mathbf{x}) = f(\mathbf{x}) + \log p(\mathbf{x}|\mathbf{w}) + \log p(\mathbf{w}). \tag{1.58}$$

For training purposes, we can neglect the constant $f(\mathbf{x})$ since it does not depend on the training variables $\mathbf{w}$. When the loss function is the logarithm of the posterior $\log p(\mathbf{w}|\mathbf{x})$ the method is called *maximum a posteriori* (MAP). The loss function then becomes

$$E(\mathbf{w}) = -\sum_n \left[\log p(\mathbf{x}_n|\mathbf{w}) + \log p(\mathbf{w})\right], \tag{1.59}$$

where the sum is over the whole dataset.

1.3.5 Tensorflow

In this thesis we use *Tensorflow* to implement our models. Tensorflow is an open source software library that was released in 2015 by Google with the purpose of making it easy for developers to design, implement and use deep learning models. On a high level, Tensorflow is a Python library where computations are expressed as a graph of *data flows*. Nodes in the graph represent mathematical operations and edges represent data that is flowing from one node to another [21]. Data in Tensorflow is represented as tensors.

Appart from Tensorflow there are a few other softwares that have been developed to implement neural networks. These include Theano, Torch, Caffe, Neon and Keras. Torch is the software developed and used at Facebook AI Research.

1.3.6 Generative models

Two major unsolved problems in the field of *artificial intelligence* (AI) are 1) data-efficiency: the ability to learn from a few data points and 2) generalization: the robustness against change of task or its context. Many machine learning models do not perform well when the input given is different from the distribution of the training set. A promising path towards overcoming these problems are generative models. They can potentially help overcome these difficulties by 1) building realistic world models that would potentially allow agents to plan before actually interacting with the world and 2) learning meaningful features of the data in an unsupervised manner. Because the features are not task specific, these could be used for downstream solutions and would be more robust and more data efficient [25].

We humans have an abstract representation of the world in our brain: we understand it is made of 3D environments; there are animals that run, fly or swim; people that talk, walk

and think; etc. We would like to endow computers with such understanding of the world. The tricky part is to develop techniques and algorithms to enable computers to learn like humans: learn in an unsupervised fashion.

As explained in the previous section, in order to build a generative model, we need to come up with a generative process. For example, the way the visual world around us was created is a combination of the laws of physics and culture. It seems rather ambitious to aspire to provide a model with the laws of physics and culture (however that could be done) and expect it to generate images that look like the world around us. Some argue that machine learning models could benefit from containing causal models, i.e. generative processes related to the process with which the training data is created [26]. A simpler approach that is usually used is to assume that there exists a probability distribution whose sampling would generate samples with the characteristics of the training set. This assumption is the very reason why probability distributions appear in generative models and this approach goes by the name *probabilistic generative modeling*. In this strategy, the task becomes to infer the probability density from the data. This usually involves the extremely challenging task of modeling a very high dimensional input data, usually specified by a full joint probability distribution.

1.3.7 Different types of generative models

1.3.7.1 GAN

Generative adversarial nets are a kind of generative model where instead of learning the probability distribution of the data, one learns to sample from it. In this framework, the generative model is pitted against an adversary: a discriminative model that learns whether a sample is from the model distribution or the data distribution. The generative process can be thought as the generative model being a counterfeiter that is trying to sneak in fake currency into the market and the discriminative model plays the role of a detective that is trying to detect the counterfeit currency. The competition in this game drives both teams to improve their methods until counterfeits are indistinguishable from the genuine currency. For a detailed explanation see [27].

1.3.7.2 VAE

In variational autoencoders, instead of doing *maximum log likelihood*, a lower bound on the evidence is maximized. The generative process consists on two steps: 1) a value $\mathbf{z}^{(i)}$ is generated from some prior distribution $p_{\theta^*}(\mathbf{z})$; 2) a value $\mathbf{x}^{(i)}$ is generated from some conditional distribution $p_{\theta^*}(\mathbf{x}|\mathbf{z})$. The ground truth generative process is hence defined

by the joint probability distribution $p_{\theta^*}(\mathbf{x}, \mathbf{z}) = p_{\theta^*}(\mathbf{x}|\mathbf{z})p_{\theta^*}(\mathbf{z})$. It is assumed that the prior $p_{\theta^*}(\mathbf{z})$ and likelihood $p_{\theta^*}(\mathbf{x}|\mathbf{z})$ come from parametric families of distributions $p_{\boldsymbol{\theta}}(\mathbf{z})$ and $p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})$. The true parameters $\boldsymbol{\theta}^*$ and the latent variables $\mathbf{z}^{(i)}$ are unknown.

The learning strategy is maximum log likelihood and if the data is *independent and identically distributed* (i.i.d.), the evidence (or marginal likelihood) of the data set $\{\mathbf{x}^{(i)}\}$ is

$$\log p_{\boldsymbol{\theta}}(\{\mathbf{x}^{(i)}\}) = \sum_i \log p_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}). \tag{1.60}$$

Because of the latent space in the model, the evidence is in general intractable because of the impossibility to marginalize over the latent variables and it is not possible to compute the integral

$$p_{\boldsymbol{\theta}}(\mathbf{x}) = \int d\mathbf{z}\ p_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z}). \tag{1.61}$$

The way around this is to introduce an inference network (or encoder network) defined over the latent variables

$$\begin{aligned} \log p_{\boldsymbol{\theta}}(\mathbf{x}) &= \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\left[\log\left(\frac{p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})p_{\boldsymbol{\theta}}(\mathbf{z})}{p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})}\frac{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\right)\right] \\ &= \underbrace{\mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\left[\log p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})\right] - D_{KL}\left[q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})||p_{\boldsymbol{\theta}}(\mathbf{z})\right]}_{ELBO} + D_{KL}\left[q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})||p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})\right], \end{aligned} \tag{1.62}$$

where the *Kulback-Leibler divergence* is defined as $D_{KL}\left[q||p\right] = \mathbb{E}_q\left[\log\left(\frac{q}{p}\right)\right]$ (the expectation value of $\log\left(\frac{q}{p}\right)$ with respect to the probability density q). The last term is intractable because it involves the computation of the evidence $p_{\boldsymbol{\theta}}(\mathbf{x})$, contained in $p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x}) = p_{\boldsymbol{\theta}}(\mathbf{z}, \mathbf{x})/p_{\boldsymbol{\theta}}(\mathbf{x})$. On the other hand $D_{KL}\left[q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})||p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})\right] \geq 0$, and so we can define an *evidence lower bound*

$$ELBO = \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\left[\log p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})\right] - D_{KL}\left[q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})||p_{\boldsymbol{\theta}}(\mathbf{z})\right]. \tag{1.63}$$

A simple choice of $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$ and $p_{\boldsymbol{\theta}}(\mathbf{z})$ is usually made to make the second term of this expression tractable. The integral in the first term called *expected reconstruction error*, cannot be analytically computed and so it needs to be estimated by sampling $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$ and the reparametrization trick is needed to backpropagate through this sampling node of the network [28].

Recall from Eq. (1.62) that

$$ELBO = \log p_{\boldsymbol{\theta}}(\mathbf{x}) - D_{KL}\left[q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})||p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})\right]. \tag{1.64}$$

Therefore, the maximization of *ELBO* concurrently maximizes the evidence and minimizes

the KL divergence between the approximate posterior and the posterior of the model. Hence one gets an approximate posterior $q_{\phi}(\mathbf{z}|\mathbf{x})$ as a byproduct of the optimization.

1.3.7.3 Autoregressive models

These models do not use latent variables. They model the joint probability by decomposing it into a product of conditionals, using the probability chain rule according to a fixed ordering over dimensions

$$p_{\theta}(\mathbf{x}) = \prod_{i} p_{\theta}(x_i|x_{i-1}, ..., x_i). \tag{1.65}$$

This form of the probability density simplifies the evaluation and sampling of the log-likelihood [29, 30].

1.3.7.4 Flow models

The basic idea behind the invention of flow models is that a good representation is one in which the data has a distribution that is easy to model [31]. In this kind of generative models, a transformation is learned from the data to a latent space $f_{\theta}(\mathbf{x}) = \mathbf{z}$, such that the resulting distribution takes a simple form, e.g. a factorized distribution. The training strategy is maximum log likelihood. In order to obtain the expression of the probability of the data, the change of variables formula is used

$$p_{\theta}(\mathbf{x}) = p_{\theta}(f_{\theta}(\mathbf{x})) \left| \det \frac{\partial f_{\theta}(\mathbf{x})}{\partial \mathbf{x}} \right|, \tag{1.66}$$

where $\frac{\partial f_{\theta}(\mathbf{x})}{\partial \mathbf{x}}$ is the Jacobian of the transformation. The mapping $f_{\theta}(\mathbf{x})$ needs to be invertible to allow sampling as follows

$$\begin{aligned} \mathbf{z} &\sim p_{\theta}(\mathbf{z}), \\ \mathbf{x} &= f_{\theta}^{-1}(\mathbf{z}). \end{aligned} \tag{1.67}$$

The challenging part in this kind of models is the task of constraining $f_{\theta}(\mathbf{x})$ to make the training computationally affordable but with the capacity to learn complex transformations.

2

Infinite lattices of vortex molecules in Rabi-coupled condensates

2.1 Introduction

In superfluids and superconductors, vortices form simple, usually triangular lattices. Tkachenko [32] proved that for an infinite system of point vortices the triangular lattice has the lowest energy, and a numerical search of up to 11 vortices per unit cell within the same model found no other stable configurations [33]. Kleiner *et al.* [34] showed that in the opposite limit of very large vortices (relative to separation), the infinite lattice orders in a triangular geometry (they show that this state has a lower energy than the square lattice that Abrikosov had erroneously suggested as a ground state [5]). Brandt later showed that the stability of the triangular lattice persists through the entire range of vortex sizes [35]. Are more complicated crystal structures possible? Superfluids with multicomponent order parameters, where vortices may form in different components, provide one avenue. Historically, the first such superfluid was ^{3}He [36], while atomic Bose condensates with internal spin states are a second, more recent example [37].

As in the solid state, one route to more complicated structures is to decorate a crystal structure with 'molecules' made of two or more vortices. This is the situation that will concern us. In 2002 Son and Stephanov [38] predicted the existence of a *vortex molecule* in a Rabi-coupled two component condensate. Subsequent works have focused on the dynamics of a single molecule [39–41], as well as ground state properties [42, 43]. Lattices of vortex molecules in a harmonic trap were studied in Refs.[44, 45].

To the best of our knowledge, vortex molecules have not been experimentally observed. There is a different object, sometimes also called vortex molecule [46], which has recently been observed in the polar phase of superfluid ^{3}He [47]. The observed object is made of two half-quantum vortices (π winding of the phase), whereas the vortex molecule we are concerned with is made of two integer vortices (2π winding of the phase). They share the feature that the vortices are linked by a domain wall which leads to confinement of the pair,

although the repulsion that balances the tension in the domain wall has a different origin, as we explain in Section 2.1.2.

Infinite vortex lattices have previously been studied both within the Lowest Landau Level (LLL) approximation and beyond [48, 49], both for single component [5, 34, 50] as well as multicomponent condensates [51–54]. For further work on vortex lattices, see the reviews [9, 49].

This chapter concerns the structure of *infinite* arrays of vortex molecules in 2D. To orient our discussion, the remainder of the introduction introduces the theoretical model and describes the physics of a single vortex molecule, before we move on to the case of a lattice.

2.1.1 Hamiltonian

We consider an infinitely extended, rotating two component spinor Bose-Einstein Condensate in 2D. In equilibrium, the thermodynamic quantity to minimize is the free energy (or energy at $T = 0$) in the rotating frame. In the presence of an AC field that gives rise to Rabi oscillations, the relevant low energy Hamiltonian is $H = H_0 + H_{\text{int}} + H_{\text{Rabi}}$, where

$$\begin{aligned} H_0 &= \sum_\sigma \int d\mathbf{r}\, \Psi_\sigma^\dagger(\mathbf{r}) \left[\frac{\mathbf{p}^2}{2} + \frac{\omega^2 r^2}{2} - \mathbf{\Omega}\cdot\mathbf{L} \right] \Psi_\sigma(\mathbf{r}) \\ &= \sum_\sigma \int d\mathbf{r}\, \Psi_\sigma^\dagger(\mathbf{r}) \left[\frac{(\mathbf{p}-\mathbf{A})^2}{2} + \frac{\omega_{\text{eff}}^2 r^2}{2} \right] \Psi_\sigma(\mathbf{r}), \end{aligned} \tag{2.1a}$$

$$H_{\text{int}} = \sum_{\sigma_1\sigma_2} \frac{g_{\sigma_1\sigma_2}}{2} \int d\mathbf{r}\, \Psi_{\sigma_1}^\dagger(\mathbf{r}) \Psi_{\sigma_2}^\dagger(\mathbf{r}) \Psi_{\sigma_2}(\mathbf{r}) \Psi_{\sigma_1}(\mathbf{r}), \tag{2.1b}$$

$$H_{\text{Rabi}} = -\Omega_R \int d\mathbf{r} \left[\Psi_a^\dagger(\mathbf{r}) \Psi_b(\mathbf{r}) + \Psi_b^\dagger(\mathbf{r}) \Psi_a(\mathbf{r}) \right]. \tag{2.1c}$$

($\hbar = m = 1$) Here, the operators $\Psi_\sigma^\dagger(\mathbf{r})$ create bosons at position $\mathbf{r}$ with spin σ, $\mathbf{A} \equiv \mathbf{\Omega} \times \mathbf{r}$, $\omega_{\text{eff}} \equiv \sqrt{\omega^2 - \Omega^2}$, ω is the harmonic trap frequency, $\mathbf{\Omega} \equiv \Omega\hat{z}$ is the angular velocity of the trap, $\mathbf{L}$ is the angular momentum operator. The dimensionless couplings $g_{\sigma_1\sigma_2} > 0$ are the strength of the hyperfine state dependent interatomic contact interactions and Ω_{R} is the Rabi frequency. The external electromagnetic field is introduced in the dipole approximation through H_{Rabi} [55].

2.1.2 Single-vortex molecules

In their seminal work, Son and Stephanov predicted that in a Rabi-coupled 3D two component BEC, there should exist a domain wall of the relative phase of the two compo-

nents –a domain wall inside which the relative phase changes by 2π [38]. This would be bound by a closed vortex line. Furthermore, they argued that the external field would work as a confinement mechanism for vortices of different components.

In this section we give qualitative arguments to explain why in 2D a pair of such vortices are confined in a *vortex molecule*. Each member of the pair is in a different component, therefore there are no logarithmic interactions within the molecule. In the mean field treatment discussed in Section 2.2, $H_{\rm int}$ and $H_{\rm Rabi}$ give rise to contributions

$$E_{\rm int} = \frac{g}{2}\int d\mathbf{r}\ \left[\rho_a(\mathbf{r}) + \rho_b(\mathbf{r})\right]^2 \ \text{and} \tag{2.2}$$

$$E_{\rm Rabi} = -2\Omega_{\rm R}\int d\mathbf{r}\ \sqrt{\rho_a(\mathbf{r})\rho_b(\mathbf{r})}\cos(\theta_a(\mathbf{r}) - \theta_b(\mathbf{r})). \tag{2.3}$$

Here we have set $g \equiv g_{aa} = g_{bb} = g_{ab}$ for simplicity, and have introduced the amplitude-phase (Madelung) representation $\psi_a(\mathbf{r}) = \sqrt{\rho_a(\mathbf{r})}e^{i\theta(\mathbf{r})}$.

For $g > 0$, $E_{\rm int}$ favors configurations where densities of different components don't overlap. $E_{\rm Rabi}$ favors alignment of the phases of the two components. Now let's think of a pair of vortices, one in each component. In the absence of the external field, there is no energy cost for having the phases misaligned:

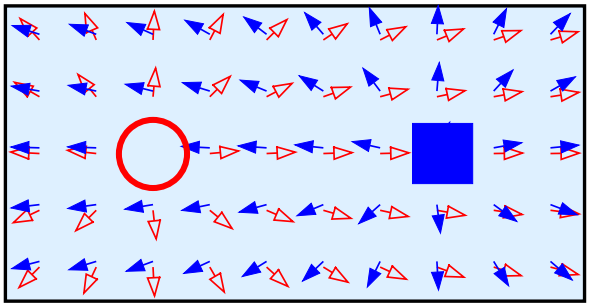


Figure 2.1: Schematic view of a pair of free vortices, one in each component, in the absence of an external field. There is no energy cost for having phases misaligned and the winding is smooth to minimize the kinetic energy. The two vortices are in different components, therefore there are no logarithmic interactions within the pair.

The term $E_{\rm Rabi}$ is minimized by full alignment of the phases. This is achieved if the two vortices overlap completely. On the other hand, overlapping vortices have an increased $E_{\rm int}$ relative to nonoverlapping vortices. Thus there is a competition between $E_{\rm Rabi}$ and $E_{\rm int}$, which have typical magnitudes per particle of $\Omega_{\rm R}$ and gn, where n is the bulk density of the two components (assumed equal).

Between the limits $\Omega_{\rm R} \ll gn$ and $\Omega_{\rm R} \gg gn$ the optimal arrangement will be a configuration where vortices are neither overlapping, nor too far separated.

Hence in the presence of vortices, which is ensured by the rotation of the trap (above

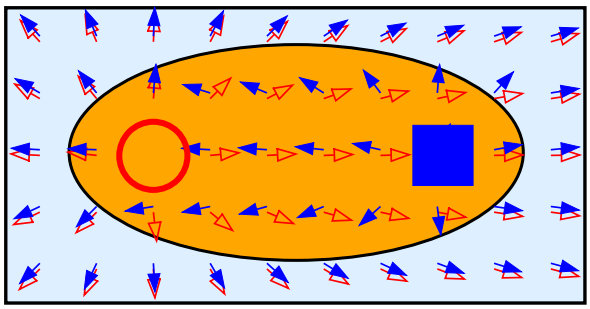


Figure 2.2: Vortex pair in the presence of the external field. Because of the energy cost for misaligning phases Eq. (2.3), the misalignment region (in orange) is confined in a region. Out of this region the phases are aligned. The two vortices are in different components, therefore there are no logarithmic interactions within the molecule.

some critical angular velocity Ω_c), the external field Ω_R works as an inter-component *confinment mechanism* for vortices, giving rise to vortex molecules. The size of the molecule (expressed in terms of the healing length $\xi = (4gn)^{-1/2}$) is a function of the ratio gn/Ω_R, but there appears to be no simple argument for this relationship. In ^{3}He-A the situation is different: the equilibrium separation of a molecule arises from a balance between tension in the domain wall and the logarithmic repulsion of the vortices [46]. In our setting the vortices are in different components, and the repulsion arises from E_{int}. Thus there is not a simple 'phase only' description of the molecule.

2.2 Gross-Pitaevskii theory

We want to study the ground state properties of the above mentioned Spinor BEC. Gross-Pitaevskii theory describes the properties of the condensate at $T = 0$. The approach used in this theory is variational, i.e. we have an interacting Hamiltonian whose exact ground state we don't know and we use our experimental knowledge about the existence of a condensate at $T = 0$, to guess the ground state wave function.

In a two level system with off-diagonal coupling, the eigenstates are superpositions of the eigenstates of the Hamiltonian in the absence of coupling [43, 56, 57]. The off-diagonal coupling mediated by the external electromagnetic field Eq. (2.1c), motivates an ansatz where there is a N particle condensate in a state which is a superposition of two hyperfine states $\{|\phi_a\rangle, |\phi_b\rangle\}$:

$$|\Psi\rangle = \frac{1}{(N!)^{1/2}} \left(\sum_\sigma \int d\mathbf{r}\, \phi^*_\sigma(\mathbf{r}) \Psi^\dagger_\sigma(\mathbf{r}) \right)^N |0\rangle. \tag{2.4}$$

We then calculate the expectation value of the Hamiltonian in this ansatz state. Defining

the condensate wave-function $\psi_\sigma(\mathbf{r}) = \sqrt{N}\phi_\sigma(\mathbf{r})$ and using that $N(N-1) \sim N^2$, the final expression to be minimized is:

$$
\begin{aligned}
E(\psi_a, \psi_b) = \sum_\sigma \int d\mathbf{r}\, \psi^*_\sigma(\mathbf{r}) \left[\frac{(\mathbf{p} - \mathbf{A})^2}{2} + \frac{\omega^2_{\text{eff}} r^2}{2} \right] \psi_\sigma(\mathbf{r}) \\
+ \sum_{\sigma_1\sigma_2} \frac{g_{\sigma_1\sigma_2}}{2} \int d\mathbf{r}\, |\psi_{\sigma_1}(\mathbf{r})|^2 |\psi_{\sigma_2}(\mathbf{r})|^2 \\
- \Omega_{\text{R}} \int d\mathbf{r}\, [\psi^*_a(\mathbf{r})\psi_b(\mathbf{r}) + \psi^*_b(\mathbf{r})\psi_a(\mathbf{r})] \,. \qquad (2.5)
\end{aligned}
$$

2.2.1 Infinite lattices

In order to study the infinite lattice, we choose $\omega_{\text{eff}} = 0$. It is the value of the effective trapping potential ω_{eff} that governs the "envelope" modulation of the condensate density. In the absence of an effective trapping potential, the only modulation is the one due to the presence of a vector potential $\mathbf{A}$. Hence, $\omega_{\text{eff}} = 0$ corresponds to having a spatially extended condensate, where there are no boundary effects and an ideal vortex lattice is expected to be found in the ground state [50]. We stress that periodicity – which we assume from now on – is not *a priori* obvious.

It is important to comment on the difference between the *unit cell* of the lattice and what we call the *computational unit cell*. The computational unit cell is the system in which we do the calculations: for numerical simplicity we choose a rectangular system. Note that this does not imply that the unit cell (the smallest portion of the system that repeats in the infinite periodic system) should be rectangular. The easiest example that illustrates this subtlety well is the triangular lattice. The unit cell is a rhombus and contains one vortex. On the other hand, if one is restricted to have a rectangular computational unit cell, this would be a rectangle with aspect ratio $\mathcal{R} = 1/\sqrt{3}$ (or $\sqrt{3}$) containing two vortices. Both unit cells reproduce the same infinite lattice.

What conditions should be imposed at the boundary of the computational unit cell? It is most natural to require gauge-invariant quantities to be periodic under some set of translations. In order to fully determine the boundary condition fulfilling the aforementioned condition, it is necessary and sufficient to require periodicity of densities, velocities and pseudo-spin: $\rho_\sigma, \mathbf{v}_\sigma$ and $\mathbf{S}$ (Appendix A.1). This leads to the boundary conditions [51, 58]:

$$
\begin{aligned}
&\psi_\sigma(x + L_x, y) = e^{i\Omega L_x y} \psi_\sigma(x, y), \\
&\psi_\sigma(x, y + L_y) = e^{-i\Omega L_y x} \psi_\sigma(x, y), \;\; \sigma = a, b. \qquad (2.6)
\end{aligned}
$$

Here L_x and L_y are the dimensions of our rectangular computational unit cell. In order to have a consistent theory, the angular velocity of the trap can only take a discrete set of values (Appendix A.2):

$$\Omega = \pi n_v, \tag{2.7}$$

where $n_v = N_v/L_xL_y$ is the vortex density in the computational unit cell.

Recently Mingarelli *et al.* [59] have studied infinite vortex lattices in a two component superfluid. The boundary conditions used in this work allow for non periodic spin solutions (Appendix A.1).

2.2.2 Numerical calculations

To find the computational unit cell and the associated ground state wavefunction, we numerically minimize the discrete version of Eq. (2.5) (see Eq. (A.18)) subject to the constraint of fixed particle number for each component, with $\omega_{\text{eff}} = 0$ and using the boundary conditions Eq. (2.6). The method used is the nonlinear conjugate-gradient algorithm as implemented in SciPy [60]. As pointed out by Mingarelli *et al.*, in order to allow the vortex lattice configuration to access any lattice geometry in the minimization process, the energy Eq. (2.5) has to be minimized not only with respect to the wavefunctions but also the aspect ratio $\mathcal{R} = L_x/L_y$ [50, 51].

To find the computational unit cell in the ground state, we use the following procedure:

1. Minimize the energy for a given $A = L_xL_y$ and N_v, to find E_{min}, $\mathcal{R}_{min}$ and $\{\psi_{\sigma,min}\}$.
2. Repeat the minimization with area $2A$ and $2N_v$ vortices.
3. If the energy has doubled and $\mathcal{R}_{min}$ has doubled (and halved – note that in general there are several $\mathcal{R}_{min}$ corresponding to one same configuration, at least $\mathcal{R}_{min}$ and $1/\mathcal{R}_{min}$), we can infer that the unit cell contains N_v vortices and its aspect ratio is $\mathcal{R}_{min}$. If either E_{min} was not doubled or $\mathcal{R}_{min}$ was not doubled and halved, we keep increasing the area and the number of vortices until the *doubling-halving criterion* has been fulfilled.
4. We repeat the same protocol for several starting N_v. We pick the solution with smallest energy density that fulfills the doubling-halving criterion.

Step 3 follows from the fact that by stacking unit cells together, one should be able to reproduce the infinite lattice. Since $\mathcal{R} = L_x/L_y$, $\mathcal{R}_{min}$ should be doubled if we stack two unit cells horizontally, and halved if we stack them vertically. The judgment of the fulfillment

of the doubling-halving criterion, takes into account the integration error Eq. (A.16). This protocol does not ensure one to find the true ground state. Note that even if we find a choice of N_v that fulfills the doubling-halving criterion, the possibility always exists that there could be a larger N_v with a lower energy density.

Note that when $\Omega_R = 0$ and $g_{aa} = g_{bb} = g_{ab}$, the energy Eq. (2.5) is invariant under unitary transformation:

$$\begin{pmatrix}\psi_a\\ \psi_b\end{pmatrix} \to U\begin{pmatrix}\psi_a\\ \psi_b\end{pmatrix},\ U \text{ unitary.} \tag{2.8}$$

As a consequence there is a continuous manifold – in fact a sphere – of ground states related by unitary transformation.

When $\Omega_R \neq 0$ and $g_{aa} = g_{bb} = g_{ab}$, the symmetry of the energy Eq. (2.5) is lowered, but it still is invariant under rotations of the spinor in the yz plane (note that E_{Rabi} is the integral of S_x Eq. (2.3)):

$$\begin{pmatrix}\psi_a\\ \psi_b\end{pmatrix} \to e^{i(\theta/2)\sigma_x}\begin{pmatrix}\psi_a\\ \psi_b\end{pmatrix}. \tag{2.9}$$

While the densities in each component ρ_σ are not invariant, the total density is. In terms of $\rho_{a,b}(\mathbf{r})$, therefore, one can find several vortex lattice geometries in the ground state.

For $\Omega_R \gg g_{ab}n$ we recover the behaviour of a scalar condensate in the state $\psi_a(\mathbf{r}) = \psi_b(\mathbf{r})$: the two vortices of each molecule (see Fig. 2.2), overlap. Thus, we recover the triangular lattice geometry, independent of the value of $\alpha \equiv g_{ab}/\sqrt{g_{aa}g_{bb}}$. We now turn to the finite Ω_R behavior. From now on we choose the values $g_{aa} = g_{bb} \equiv g = 0.125$ and $\mu_a = \mu_b \equiv \mu = 12.5n_v$, and explore three different values of α. From the Euler-Lagrange equations one can deduce that in this case the bulk densities are $n = \frac{\mu+\Omega_R}{g+g_{ab}}$ and the healing lengths $\xi = \frac{1}{\sqrt{2(\mu+\Omega_R)}}$. The choice of ξ ensures that we are both away from the LLL and point-vortex limits.

2.2.2.1 $\alpha = 1$

We begin with the case of zero Rabi field: $\Omega_R = 0$ (Fig. 2.3(a)). For this case we find that the computational unit cell contains four vortices in each of the components (the unit cell has two) and the vortex lattice is composed of two intertwined rectangular sublattices. Note that as explained above in Eq. (2.8), this is only one lattice of an infinite degenerate set. The two sublattices give rise to an overall (neglecting the two different flavors) triangular lattice $\mathcal{R} = 1/\sqrt{3}$. Our finding agrees with [53], even though we are away from the LLL limit. It would be interesting to study this problem in the opposite limit of very small healing length [61], to see whether the triangular lattice is the ground state in this case too.

Here it is interesting to note that although the computational unit cell of the density

contains two vortices of each component, the pseudo-spin has a period that is twice as large, that is why the overall computational unit cell has four vortices.

We then include Ω_R, which gives rise to molecules. In Fig. 2.4(a) we see that neighbouring molecules tend to antialign. On the other hand, we find a continuous degeneracy with respect to the alignment direction, i.e. there is no preferred direction.

2.2.2.2 $\alpha = 0.5$

For $\Omega_R = 0$ we find a square chess board configuration of vortices with four vortices per component (Fig. 2.3(b)). This result agrees qualitatively with [53].

For $\Omega_R \neq 0$, the square chess board gets distorted giving rise to a rectangular chess board (Fig. 2.4(b)). Again molecules like to align or antialign. In this case, there is a preferred direction of alignment parallel to one of the Cartesian axes.

2.2.2.3 $\alpha = 0.2$

For $\Omega_R = 0$ we find two intertwined triangular lattices with six vortices per component in the computational unit cell (Fig. 2.3(c)). This also agrees with [53].

When we include Ω_R, the result is qualitatively similar to the result obtained for $\alpha = 0.5$. The geometry of the lattice differs from the one obtained in [44] in the center of the trap. This may be due to either differences in the parameters used, because the computational unit cell is very large (we tried up to 16 vortices per component) or because for the size of the trap (relative to the healing length) used in [44], is not large enough to obtain the structure of the ideal infinite system in the bulk of the trap.

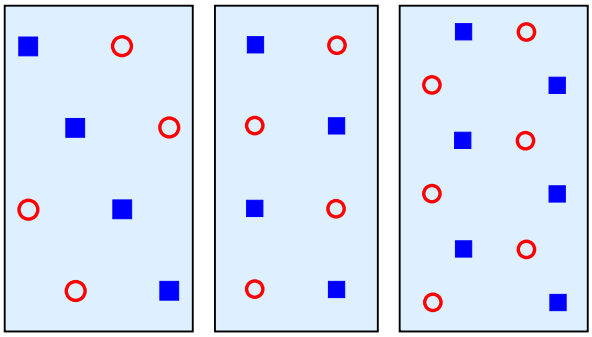
2.2.3 Orientation of the molecules

The numerical calculations show that for $\alpha = 1$ there is no preferred direction of alignment of the molecules whereas for $\alpha < 1$ there is. To understand this it is helpful to consider the configuration of the pseudo-spin density $\mathbf{S} = \Psi^\dagger \boldsymbol{\sigma} \Psi/2$. An isolated molecule has pseudo-spin pointing in the $-x$ direction at its center, and in the $+x$ direction far outside. Moving out from the center, the points where the spin points in the $\pm z$ direction give the location of the vortices in the two components.

Expressing the interaction and Rabi energies in terms of the pseudo-spin density gives

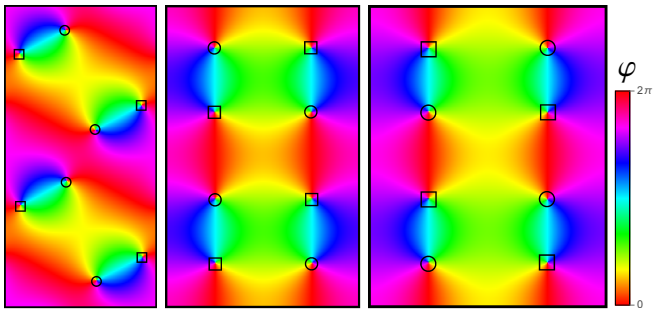
$$E_{\text{int}} = \frac{g}{4} \int d\mathbf{r} \left[4S_z^2(\mathbf{r})(1-\alpha) + \rho^2(\mathbf{r})(1+\alpha) \right] \text{ and} \tag{2.10}$$

$$E_{\text{Rabi}} = -2\Omega_R \int d\mathbf{r}\, S_x(\mathbf{r}), \tag{2.11}$$



(a) (b) (c)

Figure 2.3: Schematic view of the computational unit cell. From left to right, we show the computational unit cell for $\Omega_R/n_v = 0$ and $\alpha = 1, 0.5, 0.2$, respectively. The aspect ratios are $\mathcal{R} = 1/\sqrt{3}, 0.5, 1/\sqrt{3}$. The ground state aspect ratios are the given ones, only within integration error, i.e. there is a set of $\mathcal{R}$ within integration error that fulfill the doubling-halving criterion. Our finding agrees qualitatively with what was found in [53].



(a) (b) (c)

Figure 2.4: The colorcode shows the relative phase $\varphi = \theta_b - \theta_a$. From left to right, we show the computational unit cell for $\Omega_R/n_v = 0.5, 0.25, 0.1$ and $\alpha = 1, 0.5, 0.2$, respectively. The aspect ratios are $\mathcal{R} = 0.5, 0.64, 0.79$. The ground state aspect ratios are the given ones, only within integration error, i.e. there is a set of $\mathcal{R}$ within integration error that fulfill the doubling-halving criterion. Our finding agrees qualitatively with what was found in [44] (except $\alpha = 0.2$), in the center of the trap. Note that for $\alpha = 1$ repeated computations yield different axes along which molecules are antialigned.

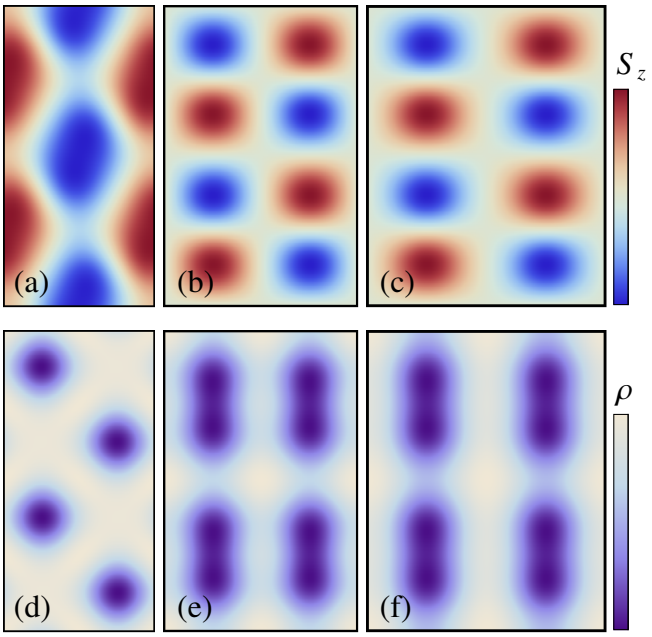


Figure 2.5: From left to right, we show the total density ρ and S_z, for $\Omega_R/n_v = 0.5, 0.25, 0.1$ and $\alpha = 1, 0.5, 0.2$, respectively. The aspect ratios are $\mathcal{R} = 0.5, 0.64, 0.79$. For $\alpha = 1$, S_z winds at constant rate around the center of each molecule, whereas for $\alpha < 1$ the winding is concentrated around the vortex cores. Consequently, for $\alpha = 1$, the density profile around the molecules is circular (at distances close enough from the center). On the other hand, for $\alpha < 1$ it is elongated.

where $\rho = \rho_a + \rho_b$. For $\alpha = 1$ the interaction energy is independent of the spin direction, so configurations that differ by global rotations of the spin around the x-axis have the same energy. Such a global rotation causes the two vortices forming the molecule to rotate about their center, explaining the numerical observation that the orientation is undetermined.

Further, an isolated molecule will have a spin that winds at a constant angular rate around the x-axis as we encircle the molecule at fixed radius. It is more natural to regard the molecule as a Skyrmion. Any modulation of the total density is circularly symmetric. For $\alpha \neq 1$ the spin configuration does not wind at a constant rate in the $y - z$ plane, and the total density is anisotropic (see Fig. 2.5).

2.3 Small molecule limit

The numerical calculations of the previous section show that neighbouring molecules are antialigned. In order to gain some insight, we develop an effective theory for the system, in the limit of very small molecules. We assume that the behavior of the lattice of molecules can be explained in terms of the kinetic energy in Eq. (2.5) only. Within this picture, g_{ab} and Ω_R take care of the shape and size of the molecule, but not its orientation.

$\Omega = \pi N_v/A$ ensures the presence of N_v vortices in each component.

2.3.1 Kinetic energy of a molecule lattice

For a system of vortices in 2D with separations $\gg \xi$, the most important contribution to the kinetic energy comes from the regions away from the vortex cores where $\nabla\sqrt{\rho_{a,b}} = 0$ and $\rho_a = \rho_b = n$ ($S_z = 0$). Using the parametrization $\Psi = \sqrt{\rho}e^{-i\chi/2}(\cos\frac{\theta}{2}e^{-i\varphi/2}, \sin\frac{\theta}{2}e^{i\varphi/2})^T$,

$$E_{\text{kin}} = \frac{n}{8}\int d\mathbf{r}\,\left[(\nabla\chi(\mathbf{r}))^2 + (\nabla\varphi(\mathbf{r}))^2\right], \tag{2.12}$$

where $\varphi = \theta_b - \theta_a$ and $\chi = -(\theta_a + \theta_b)$. Note that due to the Rabi field Eq. (2.3), $\varphi = 0$ away from the molecules and therefore the only contribution to the kinetic energy away from the molecule cores comes from the overall phase

$$E_{\text{kin}} = \frac{n}{8}\int d\mathbf{r}\,(\nabla\chi(\mathbf{r}))^2. \tag{2.13}$$

Far from a molecule, χ obeys Laplaces's equation $\nabla^2\chi = 0$, and winds by 4π as we encircle the molecule, which contains two vortices. Around an isolated molecule with no angular modulation of the density (as occurs at $\alpha = 1$, see Fig.2.5) $\chi = 2\vartheta$, where ϑ is the angular coordinate centered on the molecule. In this case, E_{kin} describes a set of point charges interacting via a 2D Coulomb interactions (see e.g. Ref. [62]). In the small molecule limit, when $\Omega_{\text{R}}/g_{ab}n$ becomes large, the kinetic energy dominates the intermolecular energy. If we consider only this contribution, then:

1. The molecules form a triangular lattice [32], as confirmed by our numerical calculations.

2. The orientation of *each* molecule is *separately undetermined*. This corresponds to a lattice of small Skyrmions, each of which may be arbitrarily rotated about the x-axis. The freedom will obviously be removed by neglected terms.

For $\alpha < 1$ the density modulation around an isolated molecule is angle dependent. Then the angular field χ of an isolated molecule will have corrections to the point charge configuration that may be described by a multipole expansion, beginning with a quadrupolar field.

The interaction of quadrupoles on a triangular lattice gives rise to a long-ranged aligning interaction between molecules that is absent for $\alpha = 1$. Recall that the size of the molecule is determined by g_{ab} and Ω_{R}, and is taken as an input to this picture.

We further assume that in the ground state the net quadrupole moment of the infinite lattice is zero $\sum_i \Delta_i^2 = 0$. As shown in Appendix A.4, with this restriction the energy per unit cell of a periodic lattice with N_{mol} molecules in the unit cell is given by:

$$V_{\text{inf}}/N_{\text{uc}} = \frac{1}{8}\text{Re}\left[\sum_{i<j} \Delta_i^2 \Delta_j^2 \mathcal{Q}(z_{ij}; \omega_1, \omega_2)\right], \tag{2.14}$$

where

$$\begin{aligned}\mathcal{Q}(z_{ij}; \omega_1, \omega_2) =& 3 \sum_{n,m} \frac{1}{(z_{ij} + n\omega_1 + m\omega_2)^4} \\ &- 4 \sum_{\substack{n,m \\ n^2+m^2>0}} \frac{1}{(n\omega_1 + m\omega_2)^4} \\ &- \sum_{k(\neq a)} \sum_{n,m} \frac{1}{(z_{ak} + n\omega_1 + m\omega_2)^4}.\end{aligned} \tag{2.15}$$

The index a corresponds to any molecule in the unit cell. $\mathcal{Q}(z_{ij}; \omega_1, \omega_2)$ is an elliptic function with periods ω_1 and ω_2. We can numerically calculate it truncating the sums. Here ω_1 and ω_2 shown in Fig. 2.6, are the vectors connecting adjacent unit cells, in complex notation [63].

2.3.2 Numerical calculations

Now that we have an expression for the energy density of the infinite lattice Eq. (2.14), we can find the lowest energy configuration in a triangular lattice with the constraint of having zero net quadrupole moment. We perform a constrained minimization using the SLSQP method implemented in SciPy. We fix the molecules Δ_i to be of unit length, so the variables are the orientations of the molecules.

Again, we assume that the infinite system is periodic and therefore we want to find the unit cell. The procedure in this case is

1. Minimize the energy Eq. (2.14) with N_{mol} unit length molecules in the unit cell.
2. Repeat the minimization, doubling the number of molecules.
3. If the energy has doubled, we can infer that the unit cell contains N_{mol} molecules.

Again, this protocol cannot guarantee that there does not exist a larger unit cell with a lower energy density.

We find that the unit cell contains two molecules. Choosing these two molecules to sit on the x axis and $\omega_2 = |\omega_2|e^{i\pi/3}$, $\mathcal{Q}(z_{12}) = |\mathcal{Q}(z_{12})|e^{-i\pi/3}$. Then $V_{\text{inf}}/N_{\text{uc}} = \frac{|\mathcal{Q}(z_{12})|}{8}\cos(2\theta_1 +$

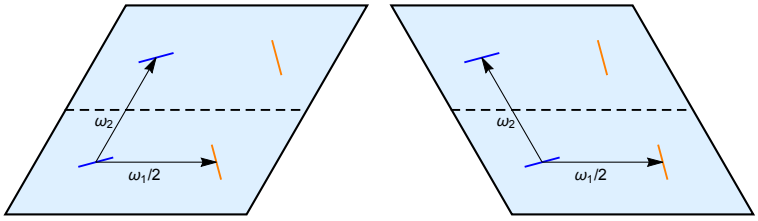


Figure 2.6: We show the two inequivalent configurations with the ground state unit cell we find.

$2\theta_2 - \pi/3$). Requiring $e^{i2\theta_1} + e^{i2\theta_2} = 0$, the ground state is $(\theta_1, \theta_2) = (\frac{\pi}{2}n + \frac{\pi}{12}) \pm (0, \frac{\pi}{2})$. There are two inequivalent infinite systems whose unit cell this is (Fig. 2.6).

2.4 Conclusions

We have calculated the ground state configuration of the infinite system of vortex molecules and found qualitative agreement with previous results in the bulk of harmonic traps. More interestingly, we have come up with an effective theory in terms of point charges and derive an expression for the interaction energy density in terms of an elliptic function $\mathcal{Q}(z_{ij}; w_1, w_2)$, in the limit of point vortices and very small molecules when the net quadrupole moment is zero. We used this expression to find the ground state configuration of an infinite system of molecules.

3

Half-vortices in spinor BECs

3.1 Introduction

Multicomponent condensates appear in various condensed matter systems. Common examples include spinor BECs, BEC mixtures and multi-orbital superconductors. The extra degrees of freedom that appear due to the vectorial nature of the order parameter, give rise to a rich variety of topological objects both in the ground state and as excitations: domain walls [64–66], Abelian [67] and non-Abelian vortices [68, 69], monopoles [70–72], 2D Skyrmions [73–75], 3D Skyrmions [76–79], vortons [80, 81], knots [82] and D-brane solitons [83]. Other works on the topic are [84, 85].

A *half-quantum vortex* (HQV) is an exotic kind of vortex that contains a discontinuity of the superfluid phase. The violation of the single-valuedness of the condensate order parameter [86] is avoided by there being a disclination of the spin vector attached to the HQV. Half vortices have been predicted to exist and experimentally observed in several condensed matter systems: superfluid ^{3}He [36, 47, 87], spin-triplet superconductors [88, 89], polariton condensates [90, 91] and BECs [74, 92, 93].

There has been previous work discussing the energy favorability of half vortices in terms of interactions. In Ref. [94] Seong-Ho Shinn *et al.* make *ansatze* for 1) a *single-quantum vortex* (SQV) at the center of a trap and 2) a pair of HQV-disclinations and analyse $\Delta E = E_{2HQV} - E_{SQV}$ as a function of interactions. They find that the pair of HQV-disclinations is favorable below certain critical value of the interaction parameters. Our work goes beyond this approach in that they impose the presence of a disclination, while we search for the ground state without this constrain.

There is also the work by Lovegrove *et al.* [95]. This is slightly different to our work in that they study HQVs in a three-component condensate, in a three-dimensional harmonic trap. It is not clear to us what the implications of these differences are, in terms of HQV favorability. They find that HQVs are stable over the entire range of parameter space that they cover.

There has also been work on the dynamics of HQVs in spin-1 BECs [96, 97]. Ji *et al.* point out the need to add a *pulsed magnetic potential* to make HQVs favorable compared to standard integer vortices.

This chapter is concerned with the nature of vortices in the ground state of a two component spinor BEC. In the Cartesian basis parametrization of the ground state spinor, we find half vortices and disclinations in certain parameter regime.

3.1.1 Hamiltonian

We consider an infinitely extended, rotating two-component spinor Bose-Einstein condensate in two dimensions. The two components correspond to two of the three hyperfine states of a spin-1 atom.

Even though there are three different hyperfine states corresponding to spin-1 atoms as we are considering here, we will only consider the two states $\sigma = \pm 1$ arguing that the state $\sigma = 0$ is frozen by the negative quadratic Zeeman coupling [93]. This is sometimes called the *easy-plane polar phase*.

The relevant low energy Hamiltonian is $H = H_0 + H_{\text{int}}$, where

$$\begin{aligned} H_0 &= \sum_\sigma \int d\mathbf{r}\, \hat{\Psi}^\dagger_\sigma(\mathbf{r}) \left[\frac{\mathbf{p}^2}{2} + \frac{\omega^2 r^2}{2} - \mathbf{\Omega}\cdot\mathbf{L} \right] \hat{\Psi}_\sigma(\mathbf{r}) \\ &= \sum_\sigma \int d\mathbf{r}\, \hat{\Psi}^\dagger_\sigma(\mathbf{r}) \left[\frac{(\mathbf{p}-\mathbf{A})^2}{2} + \frac{\omega_{\text{eff}}^2 r^2}{2} \right] \hat{\Psi}_\sigma(\mathbf{r}), \qquad (3.1a) \\ H_{\text{int}} &= \frac{W_0}{2} \sum_{\sigma_1\sigma_2} \int d\mathbf{r}\, \hat{\Psi}^\dagger_{\sigma_1}(\mathbf{r}) \hat{\Psi}^\dagger_{\sigma_2}(\mathbf{r}) \hat{\Psi}_{\sigma_2}(\mathbf{r}) \hat{\Psi}_{\sigma_1}(\mathbf{r}) \qquad (3.1b) \\ &\quad + \frac{W_2}{2} \sum_{\sigma_1\sigma_2,\sigma_1'\sigma_2'} \int d\mathbf{r}\, \mathbf{F}_{\sigma_1,\sigma_1'} \cdot \mathbf{F}_{\sigma_2,\sigma_2'} \hat{\Psi}^\dagger_{\sigma_1}(\mathbf{r}) \hat{\Psi}^\dagger_{\sigma_2}(\mathbf{r}) \hat{\Psi}_{\sigma_2'}(\mathbf{r}) \hat{\Psi}_{\sigma_1'}(\mathbf{r}) \end{aligned}$$

($\hbar = m = 1$). Here, the operators $\hat{\Psi}^\dagger_\sigma(\mathbf{r})$ create bosons at position $\mathbf{r}$ in hyperfine state σ, $\mathbf{A} \equiv \mathbf{\Omega} \times \mathbf{r}$, $\omega_{\text{eff}} \equiv \sqrt{\omega^2 - \Omega^2}$, ω is the harmonic trap frequency, $\mathbf{\Omega} \equiv \Omega\hat{z}$ is the angular velocity of the trap, $\mathbf{L}$ is the angular momentum operator and the three elements of $\mathbf{F} = \mathbf{S} + \mathbf{L}$ are the $F = 1$ matrices shown in Eq. (A.37). The dimensionless couplings are defined as $W_0 = \frac{U_0+2U_2}{3}$ and $W_2 = \frac{U_2-U_0}{3}$, where $U_0 = 4\pi a^{(0)}$ and $U_2 = 4\pi a^{(2)}$ are the scattering lengths of the s-wave channel with total angular momentum $\mathcal{F} = 0$ and $\mathcal{F} = 2$, respectively.

3.2 Gross-Pitaevskii theory

The Gross-Pitaevskii energy functional of this system is

$$E(\Psi) = \int d\mathbf{r}\left(\frac{1}{2}\partial_i\Psi^\dagger\cdot\partial_i\Psi - \mu\Psi^\dagger\cdot\Psi + \frac{W_0}{2}(\Psi^\dagger\cdot\Psi)^2 + \frac{W_2}{2}(\Psi^\dagger\mathbf{F}\Psi)^2\right), \tag{3.2}$$

where $\Psi = (\psi_1, \psi_0, \psi_{-1})^T$. We now add a quadratic Zeeman coupling term

$$E_Z = q\int d\mathbf{r}\Psi^\dagger F_z^2\Psi = q\int d\mathbf{r}\,(\rho_1 + \rho_{-1})\,, \tag{3.3}$$

where $\rho_\sigma = |\psi_\sigma|^2$. For $q < 0$, E_Z is minimised when $\rho_1 + \rho_{-1}$ is maximised. For fixed total density $\rho = \rho_1 + \rho_0 + \rho_{-1}$ this corresponds to $\rho_0 = 0$. We will assume that this term is strong enough to justify considering a two-component spinor in the ground state: $\Psi = (\psi_1, \psi_{-1})^T$. When considering a two-component spinor, E_{int} contains only density-density terms, i.e. the phases of the condensate do not enter the interactions. Hence, the model is equivalent to a BEC mixture, where no scattering events that lead to spin state transitions take place

$$E_{\text{int}} = \int d\mathbf{r}\left(\frac{W_0 + W_2}{2}\left(\rho_1^2 + \rho_{-1}^2\right) + (W_0 - W_2)\rho_1\rho_{-1}\right). \tag{3.4}$$

Note that identifying $g_{aa} = g_{bb} = W_0 + W_2$ and $g_{ab} = g_{ba} = W_0 - W_2$, this is identical to the interaction energy in Eq. (2.5). We already know everything about this model from the previous chapter.

In order to better understand the ground state and the possibility for the presence of HQVs, we move to the Cartesian basis through the following transformation [98]

$$\begin{aligned}\psi_x &= \frac{1}{\sqrt{2}}\left(-\psi_1 + \psi_{-1}\right), \\ \psi_y &= \frac{-i}{\sqrt{2}}\left(\psi_1 + \psi_{-1}\right).\end{aligned} \tag{3.5}$$

In this basis, the energy takes the form

$$E(\Upsilon) = \int d\mathbf{r}\left(\frac{1}{2}\partial_i\Upsilon^\dagger\cdot\partial_i\Upsilon - \mu\Upsilon^\dagger\cdot\Upsilon + \frac{W_0 + W_2}{2}(\Upsilon^\dagger\cdot\Upsilon)^2 - \frac{W_2}{2}|\Upsilon\cdot\Upsilon|^2\right), \tag{3.6}$$

where $\Upsilon = (\psi_x, \psi_y)^T$. For a detailed derivation of the change of basis see Appendix A.5. In the remainder of the chapter we will use the two bases interchangeably, depending on which is most convenient.

3.2.1 Ground state

In order to find the state that minimizes the interaction energy, let us go back to considering a three-component spinor Ψ in the hyperfine basis and Υ in the Cartesian basis. Let us define $\Upsilon = \mathbf{a} + i\mathbf{b}$. The average spin density for spin-1 atoms is (see Eq.(A.43))

$$\Psi^\dagger F_i \Psi = \Upsilon^\dagger (-i\epsilon_i)\Upsilon = 2(\mathbf{a}\times\mathbf{b})_i. \tag{3.7}$$

For antiferromagnetic spin interactions ($W_2 > 0$), the interactions are minimized when the spin density vanishes $\Psi^\dagger F_i \Psi = 0$, which corresponds to $\mathbf{a}$ and $\mathbf{b}$ being parallel or antiparallel. In this case, the order parameter Υ in the Cartesian basis takes the form

$$\Upsilon = \mathbf{a} + i\mathbf{b} = (a + ib)\hat{\mathbf{n}} = \sqrt{\rho}e^{i\theta}\hat{\mathbf{n}}, \tag{3.8}$$

where $\hat{\mathbf{n}}$ is a real unit vector. This is called the *polar* state. As we will see in Section 3.2.3.2, if one disregards the kinetic energy contribution to the total energy, the ground state is polar $\Upsilon = \sqrt{\rho}e^{i\theta}\hat{\mathbf{n}}$. This suggests that, at least in the case where interactions are dominant, the polar state might minimize the full energy functional Eq. (3.6).

3.2.2 Spin-gauge rotation symmetry

Interestingly, the points $(\theta + \pi, -\hat{\mathbf{n}})$ and $(\theta, \hat{\mathbf{n}})$ are equivalent in the polar state $\Upsilon = \sqrt{\rho}e^{i\theta}\hat{\mathbf{n}}$. This is sometimes called *spin-gauge rotation symmetry*. This symmetry has important implications in the physics of vortices. Concretely, it allows for the presence of half-vortices of the superfluid phase θ, in conjuction with disclinations of $\hat{\mathbf{n}}$. A disclination is a configuration where $\hat{\mathbf{n}}$ winds by π. This is because the required single-valuedness of the spinor [86] when encircling a vortex is possible even in the presence of half-integer winding of the superfluid phase, since the jump of magnitude π in the phase θ may be compensated by the corresponding sign change of the vector $\hat{\mathbf{n}}$ in a disclination. We refer to this object as a *half-quantum vortex*.

Following the usual parametrization

$$\Upsilon = \sqrt{\rho}e^{-i\chi/2}\begin{pmatrix}\cos(\theta/2)e^{-i\varphi/2}\\ \sin(\theta/2)e^{i\varphi/2}\end{pmatrix}. \tag{3.9}$$

In this parametrization, the *polar* state (where $\Psi^\dagger F_i \Psi = 0$) corresponds to $\varphi = 0$ or π. The corresponding spinors are

$$\Upsilon_0 = \sqrt{\rho}e^{-i\chi/2}\begin{pmatrix}\cos(\theta/2)\\ \sin(\theta/2)\end{pmatrix} \text{ and } \Upsilon_\pi = i\sqrt{\rho}e^{-i\chi/2}\begin{pmatrix}-\cos(\theta/2)\\ \sin(\theta/2)\end{pmatrix}. \tag{3.10}$$

In terms of the 'flag-pole' picture, these correspond to the $y = 0$ plane. Let us call this plane the *polar plane*. A HQV+disclination configuration consists of 2π winding of the pole in the polar plane, plus 2π winding of the flag, as we follow a closed path in real space. In the parametrization of the flag-pole, this corresponds to the following sequence of events as we encircle the closed path: starting from a given θ (where $\theta \in [0, \pi]$), wind to $\theta = \pi$, jump φ and χ by π, unwind to $\theta = 0$, jump φ by $-\pi$ and χ by π and wind back to the initial θ. The jumps are due to the parametrization, concretely due to $\theta \in [0, \pi]$. Both the flag and the pole undergo smooth winding, i.e. the real and imaginary parts of Υ are continuous everywhere.

3.2.3 Phase transition

We want to study the ground state of a spinor BEC under rotation in different interaction parameter regimes. To this end, it is useful to define the healing length $\xi \equiv \frac{1}{\sqrt{2\mu}}$ and dimensionless interaction parameter $\beta \equiv W_2/W_0$. We choose $W_0 = \mu/100n_v$ [where n_v is the vortex density defined in Eq. (2.7)] so that $\rho_{\text{bulk}} = 100n_v$. We start by studying the physics of varying β for fixed ξ. We work both in the hyperfine and Cartesian bases.

3.2.3.1 Hyperfine basis

In the hyperfine basis, the interaction energy density $\mathcal{E}_{\text{int}}$ takes the form

$$\mathcal{E}_{\text{int}} = \frac{1}{2\xi^2}\left[\frac{(1+\beta)}{2}\left(\rho_1^2 + \rho_{-1}^2\right) + (1-\beta)\rho_1\rho_{-1}\right]. \tag{3.11}$$

The intercomponent coupling $(1-\beta)$ changes sign at $\beta = 1$ and therefore there is a transition from repulsive to attractive interactions. Hence we already expect that for $\beta > 1$, hyperfine densities want to overlap and so vortices will overlap. For $\beta < 1$, densities repel each other and so vortices in different components will be displaced from each other. We will later show that this intuition is correct in Section 3.2.4, where we perform numerical simulations.

3.2.3.2 Cartesian basis

In the Cartesian basis the energy density takes the form

$$\mathcal{E}_{\text{int}} = \frac{1}{2\xi^2}\left[\frac{(1+\beta)}{2}\left(\Upsilon^\dagger \cdot \Upsilon\right)^2 - \frac{\beta}{2}|\Upsilon \cdot \Upsilon|^2\right]. \tag{3.12}$$

Using the usual parametrization

$$\Upsilon = \sqrt{\rho}e^{-i\chi/2}\begin{pmatrix}\cos(\theta/2)e^{-i\varphi/2}\\ \sin(\theta/2)e^{i\varphi/2}\end{pmatrix}, \tag{3.13}$$

$$\mathcal{E}_{\rm int} = \frac{\rho^2}{4\xi^2}\left[1+\frac{\beta}{2}\left(1-\cos^2\theta-\sin^2\theta\cos 2\varphi\right)\right]. \tag{3.14}$$

For $\beta > 0$, $\mathcal{E}_{\rm int}$ is minimized for $\varphi = 0, \pi$. The corresponding states are

$$\Upsilon_0 = \sqrt{\rho}e^{-i\chi/2}\begin{pmatrix}\cos(\theta/2)\\ \sin(\theta/2)\end{pmatrix} \text{ and } \Upsilon_\pi = i\sqrt{\rho}e^{-i\chi/2}\begin{pmatrix}-\cos(\theta/2)\\ \sin(\theta/2)\end{pmatrix}. \tag{3.15}$$

In the hyperfine basis and when $\beta > 1$, the interaction energy is lowered when the densities of the $\sigma = \pm 1$ components, are equal everywhere in space, i.e., $\rho_1(\mathbf{r}) = \rho_{-1}(\mathbf{r})$. The densities in the two bases are related by

$$\rho_1 = \frac{1}{2}\left(\rho_x+\rho_y+2\sqrt{\rho_x\rho_y}\sin(\theta_y-\theta_x)\right), \tag{3.16}$$

$$\rho_{-1} = \frac{1}{2}\left(\rho_x+\rho_y-2\sqrt{\rho_x\rho_y}\sin(\theta_y-\theta_x)\right). \tag{3.17}$$

Therefore $\rho_1(\mathbf{r}) = \rho_{-1}(\mathbf{r})$ if and only if $\varphi(\mathbf{r}) = \theta_y(\mathbf{r}) - \theta_x(\mathbf{r}) = 0, \pi$. Hence, for $\beta > 1$, one expects that the ground state of the system will consist of either Υ_0 or Υ_π everywhere in space.

Now what happens when $\beta < 1$? We have already understood in the hyperfine basis that $\rho_1(\mathbf{r}) \neq \rho_{-1}(\mathbf{r})$ due to the intercomponent repulsion. Therefore we cannot have a state where $\varphi(\mathbf{r}) = \theta_y(\mathbf{r}) - \theta_x(\mathbf{r}) = 0, \pi$ everywhere. An alternative is to have regions of $\varphi = 0$ and $\varphi = \pi$ separated by domain walls across which φ switches between 0 and π. We refer to this state as $\Upsilon_{\pi,0,\pi}$.

3.2.4 Numerical experiments

In order to find the ground state, we follow the method explained in Section 2.2.2 in the previous chapter. We cover a wide range of the ξ/β phase diagram. Specifically, we have looked at $\beta = 1.1$ and 0.8 in the range $\xi\sqrt{n_v} \in [0.005, 0.11]$. Also $\xi\sqrt{n_v} = 0.05$ in the range $\beta \in [0.7, 1.1]$. We have only done the full unit cell analysis explained in section 2.2.2 for $\beta = 0.8, \xi\sqrt{n_v} = 0.05$ and $\beta = 1.1, \xi\sqrt{n_v} = 0.05$. For the rest, the search has been done within the manifold of states with two vortices per component in the unit cell.

We observe the phase transition described in the previous section:

1. For $\beta > 1$, the ground state is polar, i.e., either Υ_0 or Υ_π. Fig. 3.1, corresponding to

$(\beta,\ \sqrt{n_v}\xi) = (1.1, 0.01)$, is consistent with this conclusion.

2. At $\beta < 1$, we have polar domains separated by domain walls across which φ changes from 0 to π. The ground state is what we call $\Upsilon_{\pi,0,\pi}$. Fig. 3.2, corresponding to $(\beta,\ \sqrt{n_v}\xi) = (0.8, 0.01)$, is consistent with this conclusion. Note that the phases θ_x and θ_y in Fig. 3.2 do not show the usual smooth winding around a point, but rather they feature a line discontinuity. Similarly, the profile of the partial component densities ρ_x and ρ_y shown in Fig. 3.3, do not show circular nodes but line shaped cores instead. The discontinuities in the phase are a consequence of the amplitude-phase representation of the components of the spinor Υ. In this parameterization, when both the real and imaginary parts of the wavefunction go to zero, they cross the origin in the complex plane, giving rise to a discontinuity of the phase. Both the real and imaginary parts are continuous everywhere.

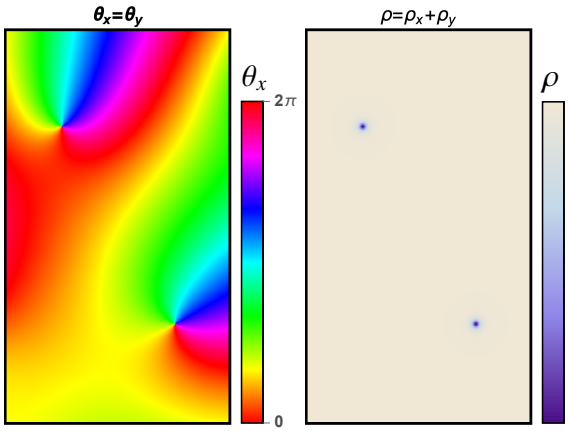


Figure 3.1: The color code in the first figure shows the phases $\theta_x = \theta_y$. The parameters with which these ground state solutions were found are $\beta = 1.1$ and $\xi\sqrt{n_v} = 0.01$.

3.2.4.1 Nature of vortices in the ground state

For $\beta > 1$, we find integer vortices both in the hyperfine and cartesian bases. On the other hand, at $\beta < 1$, in the cartesain basis, we observe the HQV+disclination configuration explained in Section 3.2.2. This can be seen in Fig. 3.4. In the hyperfine basis vortices have integer winding numbers.

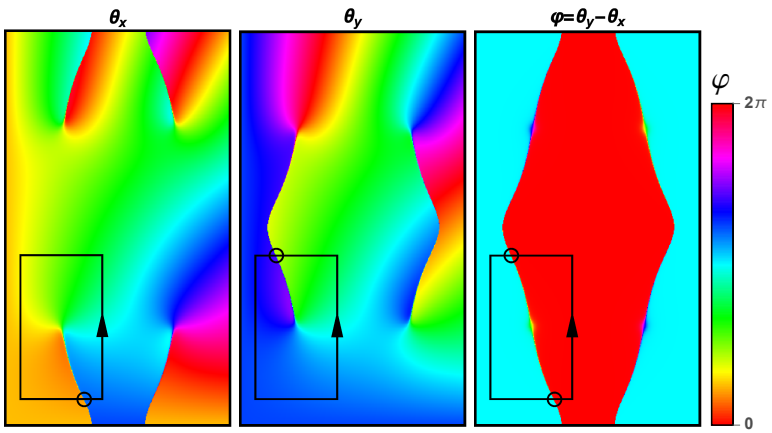


Figure 3.2: The color code in the first three figures shows the three phases θ_x, θ_y and $\varphi = \theta_y - \theta_x$. Both cartesian phases θ_x and θ_y feature a line discontinuity, unlike the usual circular vortices where the the discontinuity is pointlike. The component density modulation is accordingly not circular as shown in Fig. 3.3. The total density $\rho = \rho_x + \rho_y$ does circular modulations of the density around the vortex cores. Therefore, the relative phase $\varphi = \theta_y - \theta_x$ is also discontinuous across lines that separate regions where $\varphi = 0$ or π. The parameters with which these ground state solutions were found are $\beta = 0.8$ and $\xi\sqrt{n_v} = 0.01$.

3.3 Kinetic energy considerations

In this section we compare the energies of a single integer vortex and a pair of half-quantum vortices in the polar state, where $\Psi^\dagger F_i \Psi = 0$. The arguments used below rely on the vortices being pointlike and the core contribution being of no importance for the discussion of energy favorability. A detailed derivation of the kinetic energy of a two-dimensional system of point vortices is shown in Appendix A.6. We use the results obtained in this appendix throughout this section.

The possibility of having two different kinds of vortices in the polar phase naturally leads us to ask the question: which kind is energetically favorable? In this section, we compare the kinetic energy of a single integer vortex with that of a pair of HQV+disclinations.

As explained in 3.2.2, the polar phase corresponds to the $y = 0$ plane. We can forget about the sphere where the 'flag-pole' lives and consider the circle in the polar plane only. Then we can parametrize it by a single angle $\theta \in [0, 2\pi]$. In this case a disclination corresponds to smooth winding of θ by 2π.

For the sake of simplicity, let us forget about the previous parametrization involving half angles, and consider $\Upsilon = e^{i\chi}(\cos\theta, \sin\theta)$. In this new parametrization, a HQV corresponds to π winding of χ and a disclination corresponds to π winding of θ.

In the polar phase, the energy functional takes the simple form (neglecting the constant

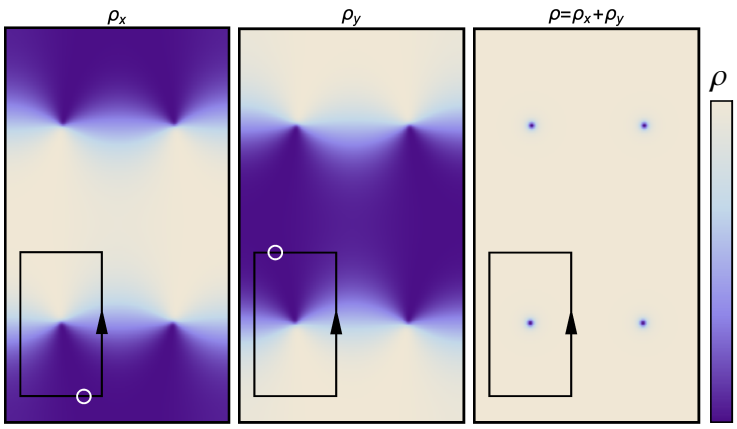


Figure 3.3: We show the component densities ρ_x, ρ_y and the total density $\rho = \rho_x + \rho_y$. Consistent with the line discontinuities shown in Fig. 3.2, the component density modulation is elongated and it is identically zero at the phase discontinuity. The total density contains circularly symmetric density modulations. These pointlike nodes are located at the intersection of the component nodal lines.

interaction term and the core contributions)

$$E(\chi,\theta) = \frac{1}{2}\int d\mathbf{r}\left((\nabla\chi)^2 + (\nabla\theta)^2\right) \equiv E_\chi + E_\theta. \tag{3.18}$$

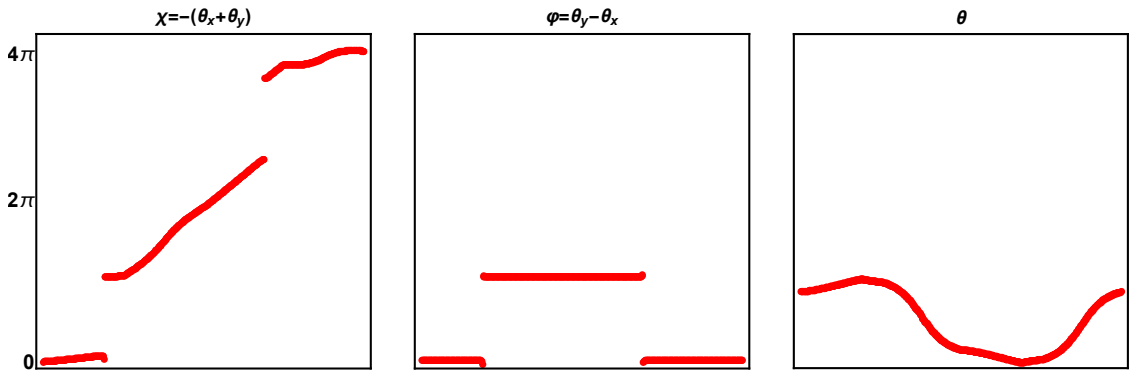
3.3.1 Single integer vortex

From solving the Euler-Lagrange equations in the rotating frame (or equivalently minimising the energy functional subject to fixed $\langle L_z\rangle$, accomplished using a Lagrange multiplier), the phase winds at constant rate, i.e., in polar coordinates, $\chi(r,\varphi) = m\varphi$. Note that for an integer vortex, single valuedness of the wave function is fulfilled without the need of a disclination and therefore it is energetically favorable to have $E_\theta = 0$. In a disc of radius R, neglecting the core contribution, the energy of this solution is

$$E_{1\,\mathrm{int}} = \pi m^2 \log\frac{R}{\xi}. \tag{3.19}$$

3.3.2 Pair of half vortices and disclinations

Both E_χ and E_θ take the form (see Appendix A.6)



(a) (b) (c)

Figure 3.4: We show the value of χ, φ and θ along the path that encloses the lower left ρ node (see Figs. 3.2 and 3.3). The total density nodes are located at the intersection of the component nodal lines. Both χ and φ feature two discontinuities as we enclose the core and the net winding of χ is 2π, giving rise to a HQV in the superfluid phase $\chi/2$. The phase $\theta = 2\text{Arccos}(\sqrt{\rho_x/\rho})$ is continuous. As explained in Section 3.2.2, this corresponds to a HQV+disclination configuration: starting from a given θ (where $\theta \in [0, \pi]$), wind to $\theta = \pi$, jump φ and χ by π, unwind to $\theta = 0$, jump φ by $-\pi$ and χ by π and wind back to the initial θ. The jumps are due to the parametrization, concretely due to $\theta \in [0, \pi]$. Both the flag and the pole undergo smooth winding, i.e. the real and imaginary parts of Υ are continuous everywhere.

$$
\begin{aligned}
E &= E_{\text{int}} + E_{\text{charge}}, \\
E_{\text{int}} &= -2\pi\rho_0 \sum_{i<j} n_i n_j \log|z_i - z_j|, \qquad (3.20) \\
E_{\text{charge}} &= \frac{\rho_0}{4i} \sum_{ij} n_i n_j \left[\oint_{\partial G} \frac{\log|z - z_i|}{(z - z_j)} dz - \oint_{\partial G} \frac{\log|\bar{z} - \bar{z}_i|}{(\bar{z} - \bar{z}_j)} d\bar{z} \right], \qquad (3.21)
\end{aligned}
$$

where ∂G is the circle $|z| = R/\xi$ with positive orientation. Let us now consider a pair of vortices with winding number n_1 and n_2, sitting in the real axis at positions $\pm r$. Then, as shown in Appendix A.6

$$
E = \frac{\pi\rho_0}{2}\left(n_1^2 + n_2^2\right)\log\left(\frac{R^2 - r^2}{\xi^2}\right) + \pi\rho_0 n_1 n_2 \left[\log\left(\frac{R^2 + r^2}{\xi^2}\right) - 2\log\left(\frac{2r}{\xi}\right)\right]. \qquad (3.22)
$$

The sign of the χ vortices is fixed by the external rotation and so $n_1 = n_2$. On the other hand θ vortices could have either same or opposite sign. The above expression shows that $n_1 = -n_2$ is favorable when $r < R/\sqrt{3}$ and not otherwise. Hence, setting $|n_1| = |n_2| = m/2$, the total energy $E_\chi + E_\theta$ is

$$E_{r<R/\sqrt{3}} = \frac{\pi m^2}{2} \log\left(\frac{R^2 - r^2}{\xi^2}\right), \tag{3.23}$$

$$E_{r>R/\sqrt{3}} = \frac{\pi m^2}{2} \log\left(\frac{R^2 - r^2}{\xi^2}\right) + \frac{\pi m^2}{2}\left[\log\left(\frac{R^2 + r^2}{\xi^2}\right) - 2\log\left(\frac{2r}{\xi}\right)\right]. \tag{3.24}$$

This shows that the integer vortex always has a higher energy. Hence, we conclude that in the polar state, in the logarithmic approximation and in the disc geometry, the kinetic energy of the pair of HQV+disclinations is lower than that of the single integer vortex.

3.4 Conclusion

We have studied the ground state of a two-component BEC as function of ξ and β. We have numerically found and analytically explained a quantum phase transition taking place at $\beta = 1$, which does not depend on ξ. We investigate the characteristics of vortices in both sides of the phase diagram and encounter half-vortices and disclinations when $\beta < 1$. We calculate the kinetic energy of a pair of HQVs and disclinations in a disc geometry and find that this is always lower than that of the integer vortex in the logarithmic approximation.

4

Matrix product states and the Stiefel manifold

4.1 Introduction

The idea behind the *Renormalization Group* (RG) is to reduce the number of microscopic degrees of freedom to the few that are relevant to the system at hand. This technique has given rise to a large number of numerical and analytical methods to study both classical and quantum many-body systems [99–101]. It wasn't until the invention of *density matrix renormalization group* (DMRG) [102, 103], however, that variational RG methods reached unprecedented accuracy in numerically studying strongly correlated one-dimensional quantum lattice systems at low temperature. Since the invention of DMRG in 1992 by Steve White [102], this technique has been widely used and considered to be the best method for the study of one-dimensional quantum lattices. The precision of this technique is limited by machine precision for a wide variety of model Hamiltonians [18].

Even though *matrix product states* (MPS) were originally discovered without reference to DMRG, instead being used as an independent tool for the analytical study of quantum systems, it was later realized that these two are intimately related. It turns out that the underlying DMRG variational *ansatz* is of the MPS form. For a detailed explanation of this connection see Ref. [18].

In recent work by Zauner-Stauber *et al.* [104], the authors claim to have developed a variational algorithm for MPS in the thermodynamic limit, that improves the convergence speed and precision with respect to DMRG.

In this chapter, we introduce a new method to calculate variational ground states of infinite 1D lattices, based on MPS. We find a connection between MPS in the *left canonical form* and the Stiefel manifold allowing us to perform optimization in this manifold. This is appealing because this manifold is smaller than the unconstrained manifold corresponding to the family of matrix product states with a given bond dimension D. Therefore, it opens the question of whether this method can compete with state of the art methods like DMRG.

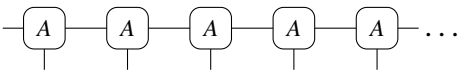
We find that our method is limited by precision. Throughout this work, we will focus on translationally invariant infinite systems.

4.2 Uniform MPS for translationally invariant infinite systems

The ground state of a translationally invariant infinite system is represented by a translationally invariant MPS in terms of matrices $A^s \in \mathbb{C}^{D\times D}$ for $s = 1, \dots d$, where d is the *physical dimension* (the number of degrees of freedom at each site) and D is the *bond dimension*

$$|\Psi(A)\rangle = \sum_{\boldsymbol{s}} \left(\dots A^{s_{n-1}} A^{s_n} A^{s_{n+1}} \dots \right) |\boldsymbol{s}\rangle \,. \tag{4.1}$$

There is one matrix A^s per lattice site, and all of them are equal, as it corresponds to a translationally invariant MPS. When working with MPS, it is convenient to use diagrams to represent the tensors. We follow the conventional notation used in the literature [18, 104]. The MPS presented above has the diagrammatic representation

$|\Psi(A)\rangle =$ 

Open bonds represent non-contracted indices and closed bonds represent contracted ones. In this case, the open bonds correspond to the physical dimension index at each s_i and the closed bonds correspond the index contracted in the matrix multiplication. We are interested in the variational ground state energy

$$E_{\text{var}} = \min_A \frac{\langle \Psi(A)|H|\Psi(A)\rangle}{\langle \Psi(A)|\Psi(A)\rangle}. \tag{4.2}$$

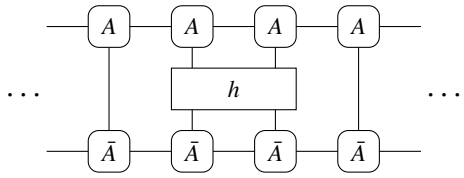
For an infinite system, this is strictly infinite, so we work with instead the energy density, which remains finite in the thermodynamic limit. In the case of a translationally invariant Hamiltonian consisting of a sum of identical, local, operators acting on neighbouring sites

$$H = \sum_j h_{j,j+1}. \tag{4.3}$$

In a translationally invariant state, the energy density is just equal to

$$e_{\text{var}} = \min_A \frac{\langle \Psi(A)|h_{j,j+1}|\Psi(A)\rangle}{\langle \Psi(A)|\Psi(A)\rangle}. \tag{4.4}$$

We represent the numerator diagrammatically as



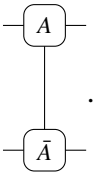
(4.5)

The transfer matrix is defined as

$$T_{ab,cd} = \sum_s A^s_{ac} \bar{A}^s_{bd}, \tag{4.6}$$

or symbolically

$T =$



.

The energy expression in Eq. (4.5) contains two infinite powers of the transfer matrix T, corresponding to the two semi-infinite halves around the sites j and $j+1$. The N^{th} power of the transfer matrix will be determined at large N by the maximal eigenvalue and the corresponding left and right eigenvectors

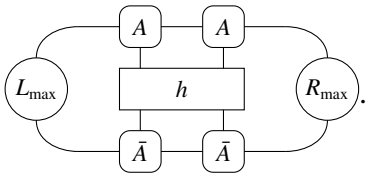
$$T^N \underset{N\to\infty}{\longrightarrow} t^N_{\max} |R_{\max})(L_{\max}|. \tag{4.7}$$

Here $(x|$ and $|x)$ denote vectorizations of a $D \times D$ matrix x in the D^2 dimensional "double layer" virtual space acted on by the transfer matrix, i.e., a $D \times D$ matrix R reshaped into a vector $|R)$ with D^2 components. ($(L_{\max}|$ and $|R_{\max})$ are the left and right eigenvectors corresponding to the maximum eigenvalue. They satisfy

$$(L_{\max}|R_{\max}) = 1. \tag{4.8}$$

If we normalize the transfer matrix so that $t_{\max} = 1$, the denominator in (4.4) is one and the numerator becomes

$e_{\text{var}} =$



. (4.9)

The key point is that $(L_{\max}|$ and $|R_{\max})$ depend on T [i.e., $|R_{\max}) \propto T^\infty(A)|v)$, where $|v)$ is any non-zero D^2 dimensional vector], and thus on the A^s. We calculate the right eigenvector

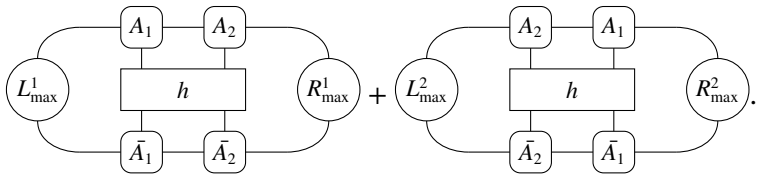
using a *power method*. This method consists on acting on an initial vector with the transfer matrix until we converge to the projector $|R_{\max})(L_{\max}|$, as shown in Eq. (4.7). Later we will see that using the right constraint for the MPS, the left principal eigenvector is the identity matrix and therefore we will not need to calculate it.

4.2.1 Multi-site unit cell

When considering an MPS with a unit cell of size two,

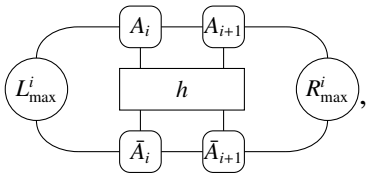
$$|\Psi(A_1, A_2)\rangle = \sum_{\boldsymbol{s}} \left(\dots A_1^{s_n} A_2^{s_{n+1}} \dots \right) |\boldsymbol{s}\rangle \,, \tag{4.10}$$

the expression for the energy density becomes

$2e_{\text{var}} =$ 

. (4.11)

For a general N-site unit cell,

$Ne_{\text{var}} = \sum_{i=1}^{N}$ 

, (4.12)

where $A_{N+1} = A_1$. Note that the principal eigenvectors depend on the pair of A tensors in each term of the sum. The N-site transfer matrix is defined as

$$T(A_1, ..., A_N)_{ab,cd} = \sum_{s_i,\dots,s_{i+N}} (A_1^{s_i} ... A_N^{s_{i+N}})_{ac} \overline{(A_1^{s_i} ... A_N^{s_{i+N}})}_{bd}. \tag{4.13}$$

The principal eigenvectors of the term $i = n$ in the sum Eq. (4.12) correspond to the transfer matrix $T(A_{n+2}, ...A_{n+N}, A_n, A_{n+1})$.

4.3 Left canonical form MPS and the Stiefel manifold

In this section we explain the connection there is between a *left canonical form* (LCF) MPS and the *Stiefel* manifold. This relation follows from: 1) the connection between the LCF MPS and completely positive (CP) maps and 2) the connection between CP maps and the Stiefel manifold.

4.3.1 Gauge degree of freedom in MPS

Let us consider the following general matrix product state

$$|\Psi(A)\rangle = \sum_{\boldsymbol{s}} \left(\dots A^{s_{n-1}} A^{s_n} A^{s_{n+1}} \dots \right) |\boldsymbol{s}\rangle \,. \tag{4.14}$$

This state is not unique in that there is a gauge degree of freedom: the MPS is invariant for any invertible matrix X of dimension $D \times D$ under

$$A^{s_i} \to A^{s_i} X, \quad A^{s_{i+1}} \to X^{-1} A^{s_{i+1}}. \tag{4.15}$$

Two special gauges that are commonly used are the *left canonical form* (LCF), defined by

$$\sum_s A^{s\dagger} A^s = \mathbb{1}, \tag{4.16}$$

and the *right canonical form* (RCF), defined by

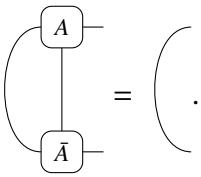
$$\sum_s A^s A^{s\dagger} = \mathbb{1}. \tag{4.17}$$

4.3.2 Left canonical form MPS and quantum channels

Any time evolution of a density matrix ρ can be seen as a map $\mathcal{E}$, that maps ρ to $\mathcal{E}(\rho)$. In quantum mechanics it is required that the map corresponding to any evolution should satisfy: linearity, complete positivity and trace preservation. According to the *Choi-Kraus representation theorem*, a map satisfies these three criteria if and only if it has a Choi-Kraus decomposition as follows [105]:

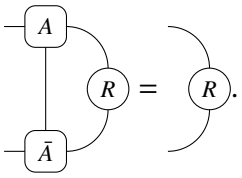
$$\mathcal{E}(\rho) = \sum_s A^s \rho A^{s\dagger}, \quad \text{where} \quad \sum_s A^{s\dagger} A^s = \mathbb{1}. \tag{4.18}$$

Such a map is often referred to as a *quantum channel* and the A^s are usually called *Kraus* operators. Note that the last condition is the definition of the *left canonical MPS* defined in Eq. (4.16). Diagrammatically, this condition is expressed as



(4.19)

The right eigenvector of the transfer matrix obeys



(4.20)

This corresponds to $\mathcal{E}(R) = R$, or that R is the fixed point density matrix of the map $\mathcal{E}$.

4.3.3 Quantum channels and the Stiefel manifold

Let us consider a unitary operator U that acts on the composite Hilbert space $\mathcal{H}_{\text{aux}} \otimes \mathcal{H}_{\text{phys}}$. Here the names auxiliary and physical are just convention. Any quantum channel can be regarded as arising from a unitary transformation in this Hilbert space

$$\mathcal{E}(\rho) = \text{tr}_{\text{phys}}\left[U\left(\rho \otimes |0\rangle\langle 0|\right)U^{\dagger}\right], \tag{4.21}$$

where ρ and $|0\rangle\langle 0|$ are density matrices in $\mathcal{H}_{\text{aux}}$ and $\mathcal{H}_{\text{phys}}$, respectively. The density matrix $|0\rangle\langle 0|$ is a pure state.

Derivation:

$$\begin{aligned}\mathcal{E}(\rho) &= \text{tr}_{\text{phys}}\left[U\left(\rho \otimes |0\rangle\langle 0|\right)U^{\dagger}\right] \\ &= \sum_s \left(\mathbb{1} \otimes \langle s|\right) U \left(\rho \otimes |0\rangle\langle 0|\right) U^{\dagger} \left(\mathbb{1} \otimes |s\rangle\right).\end{aligned} \tag{4.22}$$

We now introduce the identity in the Hilbert space H: $\mathbb{1}_H = \sum_a |a\rangle\langle a|$. Then

$$\begin{aligned}\mathcal{E}(\rho) &= \sum_s \sum_{ab} \left(|a\rangle\langle a| \otimes \langle s|\right) U \left(\rho \otimes |0\rangle\langle 0|\right) U^{\dagger} \left(|b\rangle\langle b| \otimes |s\rangle\right) \\ &= \sum_s \sum_{ab} \langle a \otimes s|\, U \left(\rho \otimes |0\rangle\langle 0|\right) U^{\dagger}\, |b \otimes s\rangle\, |a\rangle\langle b|\,.\end{aligned} \tag{4.23}$$

By further expanding $\rho = \sum_{cd} \rho_{cd} |c\rangle\langle d|$,

$$\begin{aligned}\mathcal{E}(\rho) &= \sum_s \sum_{ab} \sum_{cd} \langle a \otimes s|\, U\, |c \otimes 0\rangle\, \rho_{cd}\, \langle d \otimes 0|\, U^{\dagger}\, |b \otimes s\rangle\, |a\rangle\langle b| \\ &= \sum_s A^s \rho A^{s\dagger}, \quad \text{with} \quad A^s_{ab} = \langle a \otimes s|U|b \otimes 0\rangle\,.\end{aligned} \tag{4.24}$$

This follows from *Stinespring's theorem* [105], but also goes by the colorful name *'going to the church of the larger Hilbert space'*. This may be useful in our case as U acts in a space of dimension Dd. In fact, we don't need the full unitary. Because the density matrix $|0\rangle\langle 0|$ in the physical space is a pure state, we only need to calculate dD^2 matrix elements of the whole unitary matrix U (instead of the full d^2D^2). The unitarity of U, or equivalently the left canonical gauge of A^s, implies that we need only D orthonormal vectors in Dd dimensions, which corresponds to the *Stiefel manifold* $V_D(\mathbb{R}^{Dd})$ [or $V_D(\mathbb{C}^{Dd})$] in the complex case). The connection between CP maps and Stiefel manifolds appears in several papers on quantum control [106].

Parameterizing and optimizing in these spaces is challenging, but there is a recently-developed new Python toolbox for optimization on manifolds called Pymanopt[†]. Pymanopt aims to make it easy for users that want to use state of the art techniques for optimization on manifolds, by relying on automatic differentiation for computing gradients and Hessians, saving users time and preventing potential calculation and implementation errors. All the automatic differentiation is done behind the scenes, so that users do not need to do a large amount of set up. Pymanopt integrates with *Tensorflow*. The code where we implemented our method can be found in [107].

4.3.4 Usefulness of the connection

It is useful to optimize in the Stiefel manifold because:

1. It is a smaller manifold than the simplest alternative, which is the Euclidean manifold where all tensor elements of A are independent from each other.
2. The normalization condition is automatically satisfied.
3. We only need to calculate one of the two principal eigenvectors.

4.4 Results

Throughout this section we set $\hbar = 1$.

4.4.1 The *Affleck–Kennedy–Lieb–Tasaki* Hamiltonian

We start by testing our method on the AKLT Hamiltonian [18]

$$H = \sum_i \mathbf{S}_i \cdot \mathbf{S}_{i+1} + \frac{1}{3}(\mathbf{S}_i \cdot \mathbf{S}_{i+1})^2, \tag{4.25}$$

[†]https://pymanopt.github.io/

where i labels the sites of an infinite one-dimensional chain and $S = 1$. This model is a convenient starting point because we know its exact ground state. Furthermore, the AKLT ground state is a translational invariant infinite MPS

$$|\Psi_{AKLT}\rangle = \sum_{\boldsymbol{s}} \left(\ldots A^{s_n} \ldots \right) |\boldsymbol{s}\rangle \,, \tag{4.26}$$

where

$$A^+ = \begin{pmatrix} 0 & \sqrt{\frac{2}{3}} \\ 0 & 0 \end{pmatrix}, \quad A^0 = \begin{pmatrix} -\frac{1}{\sqrt{3}} & 0 \\ 0 & \frac{1}{\sqrt{3}} \end{pmatrix}, \quad A^- = \begin{pmatrix} 0 & 0 \\ -\sqrt{\frac{2}{3}} & 0 \end{pmatrix}. \tag{4.27}$$

We verify that our method finds the exact ground state energy density $e_0 = -2/3$, with machine precision.

4.4.2 The *Transverse Field Ising* Hamiltonian

The TFI Hamiltonian is

$$\begin{aligned} H &= -J \sum_i S_i^x S_{i+1}^x - h \sum_i S_i^z \\ &= -\frac{J}{4} \sum_i \sigma_i^x \sigma_{i+1}^x - \frac{h}{2} \sum_i \sigma_i^z, \end{aligned} \tag{4.28}$$

where again i as before and $S = 1/2$, i.e., $\{\sigma^\alpha\}$ are the Pauli matrices. The exact energy density of the ground state is [108]

$$e_0 = \frac{-h}{\pi}(1+\lambda) E\left(\theta^2\right) \text{ with } \theta^2 = \frac{4\lambda}{(1+\lambda)^2} \text{ and } \lambda = \frac{J}{2h}. \tag{4.29}$$

Here $E(\theta^2)$ is a complete elliptic integral of the second kind

$$E\left(\theta^2\right) = \int_0^{\pi/2} \left[1 - \theta^2 \sin^2 x\right]^{1/2} dx. \tag{4.30}$$

The exact ground state of the TFI Hamiltonian is not a translationally invariant MPS.

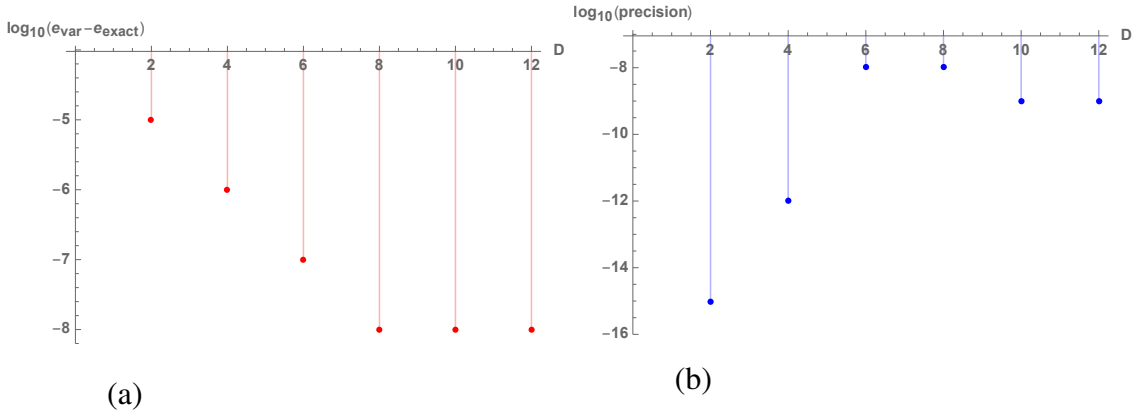
4.4.2.1 Loss of precision

Let us define two concepts we will use throughout the chapter. We define the *variational energy error* as the order of magnitude of the difference between the numerically found ground state and the exact ground state $\Delta e = e_{\text{var}} - e_{\text{exact}}$. We define the *precision* as the

order of magnitude of the difference between two converged ground states, having started from different points in the Stiefel manifold.

Even though the exact ground state of the TFI Hamiltonian is not a translationally invariant MPS, one would still expect that as the bond dimension of the MPS *ansatz* is increased, one can get arbitrarily close to the exact energy.

We use our method to try to find the ground state but we are unable to get closer than $O(10^{-8})$ to the exact energy as we increase D, i.e., the error of the variational energy $\Delta e = e_{\text{var}} - e_{\text{exact}}$ is $O(10^{-8})$. We see that this is related to the fact that as we increase the bond dimension D, the precision drops to 10^{-8}, i.e., the converged energy after the optimization depends on the starting point in the Stiefel manifold and different solutions differ by 10^{-8}. We show the variational energy density error $\Delta e = e_{\text{var}} - e_{\text{exact}}$ and the precision in Fig. 4.1.



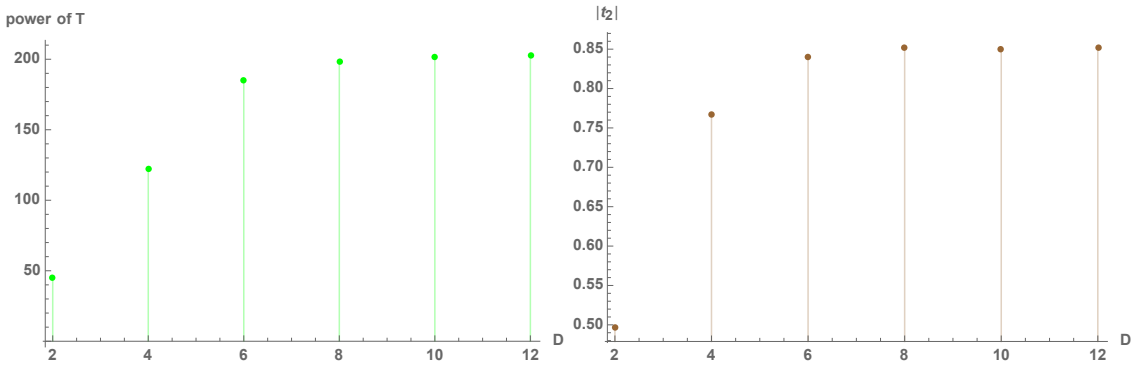
(a) (b)

Figure 4.1: We use a logarithmic scale to show (a) the error in the variational energy density and (b) the precision of the converged energies after optimization, as a function of the bond dimension D. The parameters used correspond to $h/J = 0.48$ and the unit cell contains a single site.The values shown are just the order of magnitude of the actual values, and they are an average of four experiments, i.e. four optimizations for each value of D, each starting from a different point in the Stiefel manifold.

4.4.2.2 Cause of the loss of precision

The machine precision we use is double and we use the *Conjugate Gradient* [109] method to minimize the energy. The Conjugate Gradient method is exact for quadratic forms, i.e., assuming there is no numerical error the method finds the exact solution in at most n steps, where n is the dimension of the space. The method might lose accuracy when the objective function contains very high powers of the variational parameters. This is to do with the fact that the energy landscape becomes very flat close to the minimum and therefore the norm of the gradients become very small.

The degree of our energy function Eq. (4.9) is determined by the power of the transfer matrix that goes in the right eigenvector ($|R_{\max}) = T^N(A)|v)$, where $|v)$ is any D^2-dimensional vector and N is the power needed to converge to the right eigenvector, the convergence criteria being that each element of the vector changes by less than 10^{-14}). We have observed that as the bond dimension is increased, the power of the transfer matrix goes up. We check this on the converged transfer matrix $T(A_{converged})$ and the results are shown in Fig. 4.2. Note how the precision in Fig. 4.1(b) and the power of the transfer matrix in Fig.4.2(a) are correlated.



(a) (b)

Figure 4.2: (a) We show the power of the converged transfer matrix. (b) We show the modulus of the second largest eigenvalue of the transfer matrix. The parameters used correspond to $h/J = 0.48$. The values shown are an average over four histories, i.e., four optimizations for each value of D, starting from four different points in the Stiefel manifold.

4.4.2.3 Power of the transfer matrix in the power method

The power method relies on having a principal eigenvalue with $|t_{\max}| = 1$. The number of iterations it takes to converge, is equal to the power required to remove the eigenvalue with the second largest modulus $|t_2|$, within certain tolerance. If the tolerance is 10^{-14}, then $N \approx -14/\log_{10}|t_2|$.

As shown in Fig. 4.2, our numerics show that the size of $|t_2|$ increases, i.e., $|t_1|$ and $|t_2|$ get closer, as the bond dimension D is increased. Hence the loss of precision stems from the increase of $|t_2|$.

4.4.2.4 Enlarge the unit cell to lower the power of T

A way to lower the power of the transfer matrix is to increase the size of the unit cell. The transfer matrix corresponding to the one site unit cell MPS is

$$T(A)_{ab,cd} = \sum_s A^s_{ac}\bar{A}^s_{bd}. \tag{4.31}$$

The transfer matrix corresponding to a two site MPS is

$$T(A_1, A_2)_{ab,cd} = \sum_{s_i, s_{i+1}} (A_1^{s_i} A_2^{s_{i+1}})_{ac} \overline{(A_1^{s_i} A_2^{s_{i+1}})}_{bd}. \tag{4.32}$$

Furthermore,

$$T(A_1, A_2)_{ab,cd}|R_{cd}) = T(A_1)_{ab,cd} T(A_2)_{cd,ef}|R_{ef}), \tag{4.33}$$

where repeated indices are contracted. As shown in Eq. (4.13), the N site transfer matrix is defined as

$$T(A_1, ..., A_N)_{ab,cd} = \sum_{s_i, ..., s_{i+N}} (A_1^{s_i} ... A_N^{s_{i+N}})_{ac} \overline{(A_1^{s_i} ... A_N^{s_{i+N}})}_{bd} \text{ and} \tag{4.34}$$

$$T(A, ..., A)|R) = T^N(A)|R). \tag{4.35}$$

We repeat the same optimization analysis as in the previous section, but this time we use an eight site unit cell. We find that indeed the power of the transfer matrix has decreased, but the variational energy density error shows a similar behavior as a function of the bond dimension. Probably, it is necessary to lower the power even more, by increasing the size of the unit cell, in order to decrease the energy error.

4.4.2.5 Power method problem

Besides the accuracy problem, our method is, in some cases, subject to another undesirable feature. The time taken to converge depends on $-1/\log_{10}|t_2|$, which can be very large, depending on the point in the Stiefel manifold.

The observation is that when a starting point is picked at random in the Stiefel manifold, about 50% of the times the convergence "stops" after a number of iterations, failing to converge to $|R_{\max})$. We have seen cases where $|t_{max}| - |t_2| \approx 10^{-6}$.

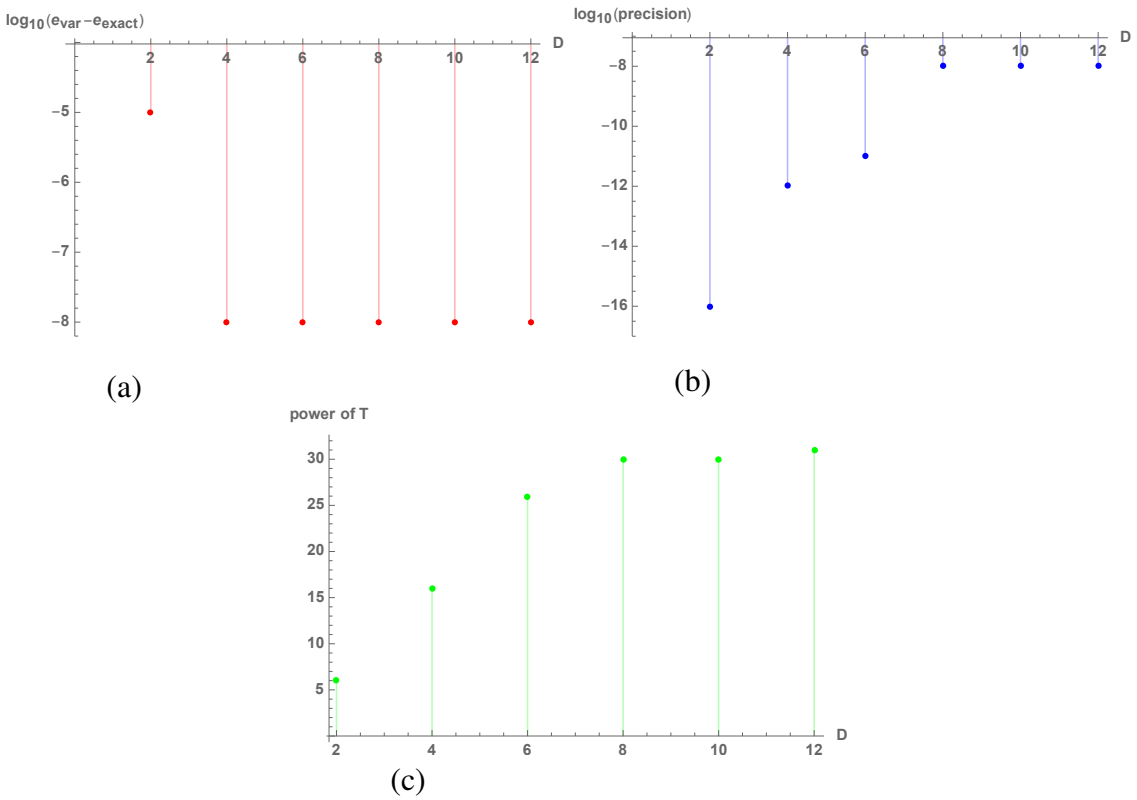


Figure 4.3: We use a logarithmic scale to show (a) the error in the variational energy density and (b) the precision of the converged energies after optimization, as a function of the bond dimension D. In (c) we show the power of the converged transfer matrix. The parameters used are $h/J = 0.48$ and the unit cell contains eight sites.The values shown are just the order of magnitude of the actual values, and they are an average of four histories, i.e. four optimizations for each value of D, starting from four different points in the Stiefel manifold.

4.5 Conclusion

We have found a connection between the *left canonical form MPS* and the *Stiefel* manifold. This allows to perform optimization in a smaller manifold than the space of matrix product states with a given bond dimension D. We have realized that our method suffers from two undesirable features. First, in order to get beyond single precision accuracy, the size of the unit cell needs to be increased and so we cannot get the acuraccy that Zauner-Stauber *et al.* [104] achieve using a translationally invariant MPS. Second, because of the presence of the power method in the variational energy expression, the convergence process gets stuck in the regions of the Stiefel manifold where $|t_2|$ is very close to one.

It remains to benchmark this method with others. There is a chance that even with the required large unit cell, our algorithm is still faster than other existing methods like

DMRG or the algorithm introduced by Zauner-Stauber *et al.* in [104].

5

Continuous matrix product states for audio synthesis

"What I cannot create, I do not understand."
– Richard Feynman

5.1 Raw audio modelling

Audio generation appears in different machine learning tasks such as music synthesis or *text-to-speech*, where the input is text and the output is speech audio. One of the reasons why it is challenging is that the dimensionality of the raw audio signal is usually a lot larger than that of the effective semantic-level signal. In speech synthesis for instance, one is typically interested in generating utterances corresponding to full sentences. At a minimum quality sampling rate of 16kHz, an average of 6,000 samples per word are generated [110].

Both music and speech are complex and highly structured. In audio signal form, different features have different timescales, ranging from milliseconds to minutes in the case of music. Because the correlations span different orders of magnitude, modeling the temporal correlations of the signals is challenging [111].

Traditionally, the high dimensionality of the raw audio modelling problem has been dealt with by compressing the audio waveforms into spectral or higher level features, and then defining generative models on these features. Examples in music generation are symbolic representations such as scores and MIDI sequences. The compression causes many of the subtleties that are crucial for the quality of sound to vanish. A way around these limitations is to model sound in the raw audio domain instead. While the digital form of audio is also lossy, the relevant information for the quality of musicality is retained.

There has been recent work on raw audio modelling using autoregressive models: AMAE [111], WaveNet [112], VRNN [113], WaveRNN [114] and SampleRNN [110].

The first is a convolutional neural network with dilated convolutions, the rest are recurrent neural networks. Beyond autoregressive models, there is WaveGlow [115] where a flow model is used and WaveGAN [116] using generative adversarial networks.

5.2 Quantum-inspired machine learning

A natural connection between quantum mechanics and machine learning is that probability distributions appear in both disciplines. Quantum-inspired machine learning is the use of quantum wave functions and quantum processes to model probability distributions and generative processes. In each case, one needs to choose the wave function and the physical process that is suitable for the problem at hand.

In raw audio modeling, the data is wave-like and quantum mechanics is a natural source of probabilistic models of wave behaviour. Hence, quantum inspired models might benefit from the *inductive bias* induced by these two characteristics: the wave-like and probabilistic nature. Furthermore, within the range of problems that exist in machine learning, one-dimensional machine learning is specially appealing for quantum many-body physicists. This is because in physics, the most powerful numerical and analytical tools have been developed to study one-dimensional systems. Therefore, there is the potential to use them to solve machine learning tasks. In this chapter, we will use *continuous matrix product states*, a numerical tool used in many-body quantum physics to handle Hilbert spaces of many-body systems, to deal with the high dimensionality of the audio data.

5.3 A quantum-inspired model for sound

We want to model raw audio using the NSynth dataset [117], which is made of four second clips of musical notes. Each note amounts to 64,000 samples and so each data point is a vector with 64,000 elements, where each element can take any real value between -1 and 1. Hence, our data lives in a very high dimensional space, which makes it unaffordable to explore brute-force: we are faced with *the curse of dimensionality*.

This is reminiscent of a problem that arises in many-body quantum optimization problems. When trying to find the variational ground state of a many-body quantum system, one has an exponentially large Hilbert space to explore. Matrix product states (MPS) serve as a tool to overcome the curse of dimensionality in this context. As explained in Section 1.3, it gives a way to parameterize the relevant corner of the Hilbert space efficiently.

The fact that MPS has proven to be a successful tool to overcome the curse of dimen-

sionality in physics suggests that it might be useful in machine learning as well. In this work, we want to explore the utility of MPS to model raw audio. On the other hand, MPS is not suitable for modeling continuous data (like raw audio), because it describes lattices of discrete degrees of freedom like spins. As explained in Section 1.2.3, there exists a generalization of MPS to systems with continuous degrees of freedom: *continuous matrix product states* (cMPS).

We will be thinking of the audio waveforms as the outcome of a sequential measurement of a continuous observable throughout the evolution of a quantum system.

We implement our model in *Tensorflow*. The codes are available in [118].

5.3.1 Physical picture of cMPS

In the following we will see that a state of the form cMPS appears in the interaction picture time evolution of a composite state of a D-level system (which we refer to as the *ancilla*), coupled to a quantum field bath. In particular we consider a D-level atom coupled to an electromagnetic field in the dipole approximation. The Hamiltonian of the composite system is

$$\begin{aligned} H &= H_a + H_b + V, \quad \text{where} \\ H_a &= \sum_n \varepsilon_n |n\rangle\langle n| , \qquad (5.1) \\ H_b &= \sum_k \omega_k b_k^\dagger b_k, \qquad (5.2) \\ V &= Ep = \sum_k \left(g_k b_k + g_k b_k^\dagger\right) \sum_{nm} p_{nm} |n\rangle\langle m| . \qquad (5.3) \end{aligned}$$

Here, $\{|n\rangle\}$ are the D eigenstates of the atom, $\{b_k\}$ are bosonic annihilation operators for each electromagnetic mode k (the quantum number k contains all the information specifying the mode), and $\{p_{nm}\}$ are the matrix elements of the dipole moment of the atom between different eigenstates. The coefficient g_k can be assumed to be real without loss of generality and it depends on details of the electromagnetic mode k, specifically the volume of the space that the modes occupy [119].

For the sake of simplicity, we will consider the case where the atom is a two-level system with energy gap Δ, and so, calling the matrix element between the two levels $p_{10} \equiv p$,

$$V = \sum_k \left(g_k b_k + g_k b_k^\dagger\right)\left(p\,|1\rangle\langle 0| + p^*\,|0\rangle\langle 1|\right). \qquad (5.4)$$

As a first step, we go to the interaction frame with respect to H_b

$$\left|\Psi^i\right\rangle = U_0^\dagger \left|\Psi\right\rangle, \quad U_0 = e^{-iH_b t}. \tag{5.5}$$

The corresponding Schrödinger equation in the interaction picture is

$$\partial_t \left|\Psi^i\right\rangle = -i\left(V_{IF} + H_a\right)\left|\Psi^i\right\rangle, \tag{5.6}$$

where the coupling in the interaction frame takes the form

$$V_{IF} = U_0^\dagger V U_0 = \sum_k \left(g_k b_k e^{-i\omega_k t} + g_k b_k^\dagger e^{i\omega_k t}\right)\left(p\left|1\right\rangle\!\left\langle 0\right| + p^*\left|0\right\rangle\!\left\langle 1\right|\right) \tag{5.7}$$

$$= \sum_k \left(g_k b_k e^{-i\omega_k t} + g_k b_k^\dagger e^{i\omega_k t}\right)\left(e^{i\Delta t}\frac{p\left|1\right\rangle\!\left\langle 0\right|}{e^{i\Delta t}} + e^{-i\Delta t}\frac{p^*\left|0\right\rangle\!\left\langle 1\right|}{e^{-i\Delta t}}\right). \tag{5.8}$$

Here, we introduced $1 = e^{i\Delta t}/e^{i\Delta t}$ to be able to perform a *rotating wave approximation* (RWA):

$$V_{IF}^{RWA} = \sum_k \left(g_k b_k^\dagger e^{-i\delta_k t} p^*\left|0\right\rangle\!\left\langle 1\right| e^{i\Delta t} + g_k b_k e^{i\delta_k t} p\left|1\right\rangle\!\left\langle 0\right| e^{-i\Delta t}\right), \tag{5.9}$$

where $\delta_k \equiv \Delta - \omega_k$ is the detuning. We define the bath operator $b(t) \equiv e^{-i\Delta t}\sum_k g_k b_k e^{i\delta_k t}$ and raising operator $R^\dagger \equiv ip\left|1\right\rangle\!\left\langle 0\right|$. The resulting Schrödinger equation takes the form

$$\partial_t \left|\Psi^i\right\rangle = \left(Rb^\dagger(t) - R^\dagger b(t) - iH_a\right)\left|\Psi^i\right\rangle. \tag{5.10}$$

The time dependence of $b(t)$ stems not only from the fact that we are in the interaction frame but also the $e^{i\Delta t}$ we introduced to perform the RWA. This is equivalently an atomic system in the Schrödinger picture, where the system is driven by these fields, which are regarded as known time-dependent operators [119].

These new operators do not follow bosonic commutation relations, instead

$$[b(t), b^\dagger(t')] = e^{-i\Delta(t-t')}\sum_k g_k^2 e^{i\delta_k(t-t')}. \tag{5.11}$$

For certain baths, $\sum_k g_k^2 e^{i\delta_k(t-t')}$ is sharply peaked at $t = t'$ [120]. Therefore, we will approximate this function with a delta function:

$$[b(t), b^\dagger(t')] = e^{-i\Delta(t-t')}\delta(t-t') = \delta(t-t'). \tag{5.12}$$

This corresponds to taking the *Markovian* limit [120]. In the remainder of the derivation,

we define the differential bath operator $dB_t \equiv b(t)dt$. We note that

$$[dB_t, dB_t^\dagger] = \underbrace{\delta(0)dt}_{1} dt = dt. \tag{5.13}$$

and so $dB_t \sim \sqrt{dt}$. This can be understood by thinking of dt as the smallest unit into which time can be devided. Then we have a discrete delta function with finite width and height, with area $\delta(0)dt = 1$.

We now consider the case where the electromagnetic field is in the vacuum state $|0\rangle$. Considering a differential step and expanding to order dt

$$\begin{aligned}\left|\Psi^i_{dt}\right\rangle &= \exp\left(RdB^\dagger_{dt} - R^\dagger dB_{dt} - iH_a dt\right)\left|\psi^i_0\right\rangle \otimes |0\rangle \\ &\approx \left(\mathbb{1} - \left(iH_a + \frac{R^\dagger R}{2}\right)dt + RdB^\dagger_{dt} + \frac{R^2}{2}dB^\dagger_{dt}dB^\dagger_{dt}\right)\left|\psi^i_0\right\rangle \otimes |0\rangle\,. \end{aligned} \tag{5.14}$$

Note that $dB_{dt}|0\rangle = 0$ and $dB_{dt}dB^\dagger_{dt}|0\rangle = dt|0\rangle$ from Eq. (5.13). Neglecting the last term, this is the first order expansion of the continuous matrix product state defined in Eq. (1.29), given we identify $Q = -iH_a - R^\dagger R/2$ and $\hat{\psi}^\dagger dt = dB^\dagger$. A more careful analysis (beyond the scope of this work) reveals that the last term $dB^\dagger dB^\dagger$ need not be kept [119]. Thus we are finally left with a *continuous matrix product state* [19, 121]

$$\left|\Psi^i_{dt}\right\rangle \approx \left(\mathbb{1} - \left(iH_a + \frac{R^\dagger R}{2}\right)dt + RdB^\dagger_{dt}\right)\left|\psi^i_0\right\rangle \otimes |0\rangle\,. \tag{5.15}$$

As a last step let us consider the time evolution of $\left|\Psi^{i'}_{dt}\right\rangle = e^{iH_a t}\left|\Psi^i_{dt}\right\rangle$ so that the model takes a more compact form. Then,

$$\left|\Psi^{i'}_{t+dt}\right\rangle = \left[\mathbb{1} - \frac{R_t^\dagger R_t}{2}dt + R_t dB^\dagger_{t+dt}\right]\left|\psi^{i'}_t\right\rangle \otimes |0\rangle\,, \tag{5.16}$$

where $R_t = e^{iH_a t}Re^{-iH_a t}$. In the remainder, we will not keep the i' index but we will still be referring to states whose time evolution is (5.16). Also we will use H to refer to H_a.

5.3.2 Balanced homodyne detection

Balanced homodyne measurement corresponds to mixing of the output field with a strong (classical) oscillator (mode a) on a balanced beam splitter, and measuring the photon

number difference between the two output fields $c = (a + b)/\sqrt{2}$ and $d = (a - b)/\sqrt{2}$:

$$\Delta n = c^\dagger c - d^\dagger d = ab^\dagger + a^\dagger b \approx \alpha b^\dagger + \alpha^* b, \tag{5.17}$$

where the last approximation follows from the operator Δn acting on the coherent state $|\alpha\rangle$, i.e., $a|\alpha\rangle \approx \alpha|\alpha\rangle$. In particular, the approximation becomes exact for the photon count divided by the oscillator amplitude in the strong oscillator limit,

$$I \equiv \lim_{|\alpha|\to\infty} \frac{c^\dagger c - d^\dagger d}{|\alpha|} = e^{i\phi} b^\dagger + e^{-i\phi} b. \tag{5.18}$$

We now discuss the effect of the operator $I = e^{i\phi} b^\dagger + e^{-i\phi} b$ being measured continuously on the output of an open quantum system described by a cMPS. As shown in Eq. (5.16),

$$\begin{aligned} |\Psi_{t+dt}\rangle &= \left[\mathbb{1} - \frac{1}{2} R_t^\dagger R_t dt + R_t \otimes dB_{t+dt}^\dagger\right] |\psi_t\rangle \otimes |0\rangle \\ &= \left[\mathbb{1} - \frac{1}{2} R_t^\dagger R_t dt + R_t \otimes (dB_{t+dt}^\dagger + e^{-i2\phi} dB_{t+dt})\right] |\psi_t\rangle \otimes |0\rangle, \end{aligned} \tag{5.19}$$

where we introduced $e^{-i2\phi} dB_{t+dt}$ so that we can introduce the operator I in the equation of motion. If we make a measurement of I at time $t + dt$, projecting the state $|\Psi_{t+dt}\rangle$ onto $|I_{t+dt}\rangle \otimes \langle I_{t+dt}|\Psi_{t+dt}\rangle$, we are left with the following state of the ancilla

$$\langle I_{t+dt}|\Psi_{t+dt}\rangle \equiv |\widetilde{\psi}_{t+dt}\rangle = \left[\mathbb{1} - \frac{1}{2} R_t^\dagger R_t dt + R_t\, e^{-i\phi} I_{t+dt} dt\right] |\psi_t\rangle \times \sqrt{\mathcal{P}(I_{t+dt})}, \tag{5.20}$$

where $\mathcal{P}(I_{t+dt}) = |\langle 0|I_{t+dt}\rangle|^2 = \sqrt{dt/2\pi}\exp(-dt I_{t+dt}^2/2)$ is the probability of measuring I_{t+dt} on the vacuum state and

$$\begin{aligned} p(I_{t+dt}) = \langle\widetilde{\psi}_{t+dt}|\widetilde{\psi}_{t+dt}\rangle &= \mathcal{P}(I_{t+dt}) \Big\{ 1 + \langle e^{-i\phi} R_t + e^{i\phi} R_t^\dagger\rangle_{\psi_t} I_{t+dt} dt \\ &\quad + \left[-1 + \left(I_{t+dt}\sqrt{dt}\right)^2\right] \langle R_t^\dagger R_t\rangle_{\psi_t} dt + \mathcal{O}(I_{t+dt} dt^2, dt^2) \Big\} \end{aligned} \tag{5.21}$$

is the probability density of obtaining I_{t+dt}. Recalling Eqs. (5.13) and (5.18), note that $I \sim 1/\sqrt{dt}$. Then $\left[-1 + \left(I_{t+dt}\sqrt{dt}\right)^2\right]$ is of order one and

$$p(I_{t+dt}) = \sqrt{\frac{dt}{2\pi}} \exp\left[-\frac{dt}{2} I_{t+dt}^2 + \langle e^{-i\phi} R_t + e^{i\phi} R_t^\dagger\rangle_{\psi_t} I_{t+dt} dt + \mathcal{O}(dt)\right]. \tag{5.22}$$

We now add a term of order dt to complete the square so that we get a Gaussian probability

density

$$
\begin{aligned}
p(I_{t+dt}) &= \sqrt{\frac{dt}{2\pi}} \exp\left[-\frac{dt}{2}\Big(I_{t+dt} - \langle e^{-i\phi}R_t + e^{i\phi}R_t^\dagger\rangle_{\psi_t}\Big)^2 + \mathcal{O}(dt)\right] \\
&\approx \sqrt{\frac{dt}{2\pi}} \exp\left[-\frac{dt}{2}\Big(I_{t+dt} - \langle e^{-i\phi}R_t + e^{i\phi}R_t^\dagger\rangle_{\psi_t}\Big)^2\right] \\
&= \sqrt{\frac{1}{2\pi\left(1/\sqrt{dt}\right)^2}} \exp\left[-\frac{\Big(I_{t+dt} - \langle e^{-i\phi}R_t + e^{i\phi}R_t^\dagger\rangle_{\psi_t}\Big)^2}{2\left(1/\sqrt{dt}\right)^2}\right].
\end{aligned} \tag{5.23}
$$

Equivalently,

$$
I_{t+dt} = \langle e^{-i\phi}R_t + e^{i\phi}R_t^\dagger\rangle_{\psi_t} + z, \text{ where } z \sim N(0, 1/dt), \tag{5.24}
$$

where $N(0, 1/dt)$ is a Gaussian distribution with zero mean variance $1/dt$. For the remainder of the chapter we fix $\phi = 0$. The conditional joint probability density for a sequence of measurements $\{I_t\}$ is

$$
\begin{aligned}
p(I_T, ..., I_1|H, R) &= \prod_{k=0}^{T-1} p(I_{k+1}|I_k, ..., I_1; H, R), \text{ where} \\
p(I_{k+1}|I_k, ..., I_1; H, R) &= \sqrt{\frac{1}{2\pi\left(1/\sqrt{dt}\right)^2}} \exp\left[-\frac{\Big(I_{k+1} - \langle R_t + R_t^\dagger\rangle_{\psi_k}\Big)^2}{2\left(1/\sqrt{dt}\right)^2}\right],
\end{aligned} \tag{5.25}
$$

where ψ_k is the state of the ancilla at time k.

5.3.3 The model

Our model generative process consists on the continuous measurement of the homodyne current I_t, on the output of an open quantum system described by a cMPS. As a refinement of the cMPS model, we include two extra variables: A and σ. The model involves the signal I_t together with a latent Hilbert space consisting of vectors $|\psi\rangle \in \mathbb{C}^D$. The signal follows the stochastic process

$$
I_{t+dt} = A\langle R_t + R_t^\dagger\rangle_t + z, \quad \text{where} \quad z \sim N(0, 1/dt). \tag{5.26}
$$

The parameter A is a real learning variable, $R_t = e^{iHt}Re^{-iHt}$ (H is real and diagonal), $R \in \mathbb{C}^{D\times D}$ is a matrix acting on the latent space and the angular brackets $\langle\cdot\rangle_t$ denote the

quantum mechanical expectation over an (unnormalized) state $|\tilde{\psi}_t\rangle$

$$\langle\cdot\rangle_t = \frac{\langle\tilde{\psi}_t|\cdot|\tilde{\psi}_t\rangle}{\langle\tilde{\psi}_t|\tilde{\psi}_t\rangle}. \tag{5.27}$$

The state $|\tilde{\psi}\rangle$ evolves according to

$$|\tilde{\psi}_{t+dt}\rangle = \left[\mathbb{1} - \frac{\sigma^2}{2}R_t^\dagger R_t dt + R_t I_{t+dt} dt\right]|\tilde{\psi}_t\rangle\,, \tag{5.28}$$

$$|\psi_{t+dt}\rangle = |\tilde{\psi}_{t+dt}\rangle \,/\, \sqrt{\langle\tilde{\psi}_{t+dt}|\tilde{\psi}_{t+dt}\rangle}. \tag{5.29}$$

The purpose of introducing the training variable A is to learn the amplitude of the signal. The amplitude is set by $A\langle R_t + R_t^\dagger\rangle_t$ in Eq. (5.26). This is done to learn R independently of the amplitude of the signals in the dataset. This way, the training of R is geared solely towards optimizing the time evolution of $|\tilde{\psi}_t\rangle$ in Eq. (5.28). The hyperparameter dt sets the strength of the term $R_t^\dagger R_t$ compared to $R_t I_{t+dt}$ (this is trivial to see by absorbing $\sqrt{dt}$ into R). In cases where we are interested in fixing dt to be the real time discretization of the data, is σ the hyperparameter in charge of this. The initial state $|\psi_0\rangle$ is learned.

The conditional joint probability density for a sequence of measurements $\{I_t\}$ is

$$\begin{aligned} p(I_T, ..., I_1|H, R, A, |\psi_0\rangle) &= \prod_{k=0}^{T-1} p(I_{k+1}|I_k, ..., I_1; H, R, A, |\psi_0\rangle), \text{ where} \\ p(I_{k+1}|I_k, ..., I_1; H, R, A, |\psi_0\rangle) &= \sqrt{\frac{dt}{2\pi}} \exp\left[-\frac{dt}{2}\left(I_{k+1} - A\langle R_k + R_k^\dagger\rangle_k\right)^2\right]. \end{aligned} \tag{5.30}$$

This constitutes an *Autoregressive Recurrent Neural Network* where the hidden state is the quantum wavefunction $|\tilde{\psi}\rangle$ and the non-linear update equation is Eq. (5.29). Aside from a few last details that will be explained in the coming sections, this probability distribution defines our *quantum-inspired model*.

5.4 Data as homodyne current

We have seen that the sequential measurement of the homodyne current on the output of an open quantum system described by a cMPS gives rise to the autoregressive probability distribution shown in Eq. (5.30). We now want to use this probability distribution to model raw audio data x_t. One obvious thing to do is to consider the raw audio to be the homodyne current, i.e. $I_t \equiv x_t$, in which case the generative model is defined as

$$p(x_T, ..., x_1|H, R, A, |\psi_0\rangle) = \prod_{k=0}^{T-1} p(x_{k+1}|x_k, ..., x_1; H, R, A, |\psi_0\rangle),$$
$$p(x_{k+1}|x_k, ..., x_1; H, R, A, |\psi_0\rangle) = \sqrt{\frac{dt}{2\pi}} \exp\left[-\frac{dt}{2}\left(x_{k+1} - A\langle R_k + R_k^\dagger\rangle_k\right)^2\right]. \tag{5.31}$$

Note that in the limit $dt \to 0$, the variance $1/dt$ diverges and the samples of this probability distribution become pure noise. Hence, this model does not have a continuous limit in the sense that as the time discretization becomes dense, the signal does not become smoother but more discontinuous.

If our training strategy is *maximum log likelihood*, the loss function of a single data point is

$$-\log p(x_T, ..., x_1|H, R, A, |\psi_0\rangle) = -\sum_{k=0}^{T-1} \log p(x_{k+1}|x_k, ..., x_1; H, R, A, |\psi_0\rangle). \tag{5.32}$$

Since dt is a hyperparameter (i.e., we do not learn it), we define the loss function as

$$\text{loss}(H, R, A, |\psi_0\rangle) = \sum_{k=0}^{T-1} \left(x_{k+1} - A\langle R_k + R_k^\dagger\rangle_k\right)^2. \tag{5.33}$$

At sampling time, the variance of the Gaussian in Eq. (5.31) is tuned by introducing a temperature parameter T (explained in Sec. 5.5.7) to optimize the quality of the samples. Therefore, dt does not influence the variance at generation time.

5.5 Time derivative of data as homodyne current: a stochastic differential equation perspective

Let us consider Eq. (5.26). Note that $z \sim N(0, q^2)$ is equivalent to $qz \sim N(0, 1)$ and therefore multiplying both sides by dt

$$I_{t+dt}dt = A\langle R_t + R_t^\dagger\rangle_t dt + d\beta_t. \tag{5.34}$$

The process β_t is Brownian motion and its independent increments have variance dt. In the limit $dt \to 0$, this equation is reminiscent of a *stochastic differential equation*

$$dI_t = \text{f}(I_t, t)dt + d\beta_t. \tag{5.35}$$

On the other hand, the left hand side of Eq. (5.34) contains the value of the stochastic process I_t whereas the left hand side of Eq. (5.35) contains the differential of the process $dI_t \equiv I_{t+dt} - I_t$. In order to rephrase our model in the language of stochastic differential equations, an option is to define the time derivative of the raw audio data to be the outcome of the homodyne current measurement, i.e. $I_t \equiv dx_t/dt$, instead of $I_t \equiv x_t$.

As explained in Ref. [122], there are several advantages of having a continuous formulation of the model, even though one always needs to discretize to perform numerical calculations. One of the main advantages is that one can use the machinery developed to numerically integrate stochastic differential equations.

Even though it is appealing to rephrase the model in terms of SDEs, later we will see that this is not always a good option, since for certain datasets, the time derivative dx_t/dt of the signals is more spiky and discontinuous than the signal x_t, which is problematic for training our model. In the remainder of the chapter, we will use both approaches. Depending on the choice, the notation will be

- If $I_t \equiv x_t$, then $x_{t+dt} = A\langle R_t + R_t^\dagger\rangle_t + z$, where $z \sim N(0, 1/dt)$.
- If $I_t \equiv dx_t/dt$, then $dx_t = A\langle R_t + R_t^\dagger\rangle_t dt + d\beta_t$.

Throughout the chapter, unless we consider it helpful, we do not specify the units of different quantities.

5.5.1 The model from an SDE perspective

In the continuous formulation of the model, the signal follows the stochastic process (Itô process)

$$dx_t = A\langle R_t + R_t^\dagger\rangle_t dt + d\beta_t. \tag{5.36}$$

Here β_t is Brownian motion with diffusion constant q, which means that the independent increments $\Delta\beta \equiv \beta_{k+1} - \beta_k$ are zero mean Gaussian random variables with variance $q\Delta t$. The state $|\tilde{\psi}\rangle$ evolves according to

$$d\,|\tilde{\psi}_t\rangle = \left[-\frac{\sigma^2}{2}R_t^\dagger R_t dt + R_t dx_t\right]|\tilde{\psi}_t\rangle\,. \tag{5.37}$$

Hence our model in Eq. (5.36) has the form of a non-linear stochastic differential equation.

5.5.2 Generalization to density matrices

To add expressivity to the model, we can consider starting from a learned density matrix ρ_0, and evolving the density matrix instead of the state $|\tilde{\psi}_t\rangle$. There is the disadvantage that $\rho \in \mathbb{C}^{D\times D}$, so it is more costly to evolve than the pure state. The equation of motion for the (unnormalized) density matrix is

$$\frac{d\tilde{\rho}_t}{dt} = \sigma^2 L(\tilde{\rho}_t) + (\tilde{\rho}_t R_t^\dagger + R_t\tilde{\rho}_t)\frac{dx_t}{dt}, \tag{5.38}$$

where $L(\cdot)$ is the Linbladian

$$L(\rho) = R_t\rho R_t^\dagger - \frac{1}{2}\left(R_t^\dagger R_t\rho + \rho R_t^\dagger R_t\right). \tag{5.39}$$

The quantum mechanical average in Eq. (5.36) then becomes

$$\langle R_t + R_t^\dagger\rangle_t = \frac{\mathrm{Tr}\left[\left(R_t + R_t^\dagger\right)\tilde{\rho}_t\right]}{\mathrm{Tr}\left[\tilde{\rho}_t\right]}. \tag{5.40}$$

5.5.3 Parameter estimation

We now have a parametric form of our model and we need to find the values of the parameters that best fit the data, given a dataset. We will use the dataset to determine those values, case by case. A classical method for SDE parameter estimation is the *maximum likelihood* method. In this method, the aim is to maximise the probability of the data, given the parameters of the model. An alternative approach is to do *maximum a posteriori* estimation, where the strategy is to maximize the *posterior* probability of the parameters, given the data. Using Bayes' theorem, we relate the posterior to the likelihood and the prior

$$\begin{aligned}\underbrace{p(\boldsymbol{\theta}|x_1,...,x_T)}_{posterior} &= \frac{\overbrace{p(x_1,...,x_T|\boldsymbol{\theta})}^{likelihood}\overbrace{p(\boldsymbol{\theta})}^{prior}}{\underbrace{p(x_1,...,x_T)}_{evidence}} \\ &\propto p(\boldsymbol{\theta})\prod_{k=0}^{T-1} p(x_{k+1}|x_k,...,x_1;\boldsymbol{\theta}).\end{aligned} \tag{5.41}$$

Here $\boldsymbol{\theta}$ are the learning variables. We do not need to include the *evidence* in the optimization because it does not depend on $\boldsymbol{\theta}$. In order to use either of the two aforementioned methods, we need to be able to evaluate the likelihood, which is in general hard. In the case of

non-linear SDEs, like our model (5.36), a typical approach is to replace the SDE with a tractable approximation. We can for example, use an SDE discretisation method.

The probability distribution of continuous processes is not normalizable, i.e. if we formally define it as

$$p(\mathcal{X}_t) = \lim_{n\to\infty} p(x_1, ..., x_n), \tag{5.42}$$

this limit tends to zero or infinity almost everywhere in the domain of the distribution. As an intuitive example, consider

$$p(\mathcal{X}_t) = \lim_{n\to\infty} \prod_{i=1}^{n} p(x_i). \tag{5.43}$$

Then, $p(\mathcal{X}_t)$ tends to zero or infinity everywhere, unless there are points where $p(x_i) = 1$ for all i. Hence, for a strictly continuous process, its likelihood is not normalizable. In order to define a finite loss function, we can consider the relative probability distribution of the process $\mathcal{X}_t$ with respect to the probability of another process that does not contain the learning parameters. It is natural to define the relative probability of the signal $\mathcal{X}_t$ with respect to the driving Brownian motion β_t. Let us call the probability measure of our model $\mathbb{P}_{\text{cMPS}}(\mathcal{X}_t)$ and the probability distribution associated with Brownian motion $\mathbb{P}_\beta(\mathcal{X}_t)$. According to the Girsanov theorem ([123]), the relative probability of $\mathbb{P}_{\text{cMPS}}(\mathcal{X}_t)$ with respect to $\mathbb{P}_\beta(\mathcal{X}_t)$ is given by the Radon-Nikodym derivative involved in changing measure from $\mathbb{P}_{\text{cMPS}}(\mathcal{X}_t)$ to $\mathbb{P}_\beta(\mathcal{X}_t)$:

$$\frac{d\mathbb{P}_{\text{cMPS}}(\mathcal{X}_t)}{d\mathbb{P}_\beta(\mathcal{X}_t)} = \exp\left(\frac{A}{q}\int \langle R_t + R_t^\dagger\rangle_t dx_t - \frac{A^2}{2q}\int \langle R_t + R_t^\dagger\rangle_t^2 dt\right), \tag{5.44}$$

where q is the diffusion constant of the Brownian motion. Our training strategy is to minimise the negative log likelihood (relative to the measure of Brownian motion), i.e., our loss function (associated to a single continuous audio signal $\mathcal{X}_t$) is minus the logarithm of Eq. (5.44):

$$\text{loss} = -\log\frac{d\mathbb{P}_{\text{cMPS}}(\mathcal{X}_t)}{d\mathbb{P}_\beta(\mathcal{X}_t)} = -A\int \langle R_t + R_t^\dagger\rangle_t dx_t + \frac{A^2}{2}\int \langle R_t + R_t^\dagger\rangle_t^2 dt. \tag{5.45}$$

Note that we removed the factor $1/q$ because it is not a learning variable.

5.5.4 Discretization

Since we cannot solve Eq. (5.36) exactly, we cannot evaluate the likelihood exactly and so we need the aid of discretization methods. We use the Euler-Maruyama integration scheme

$$|\tilde{\psi}_{t+\Delta t}\rangle = \left[\mathbb{1} - \frac{\sigma^2}{2}R_t^\dagger R_t \Delta t + R_t \Delta x_t\right]|\tilde{\psi}_t\rangle\,, \quad \text{and} \tag{5.46}$$

$$\tilde{\rho}_{t+\Delta t} = \left[\mathbb{1} - \frac{\sigma^2}{2}R_t^\dagger R_t \Delta t + R_t \Delta x_t\right]\tilde{\rho}_t\left[\mathbb{1} - \frac{\sigma^2}{2}R_t^\dagger R_t \Delta t + R_t \Delta x_t\right]^\dagger. \tag{5.47}$$

The discretization of the model SDE (5.36) is

$$\Delta x_t = A\langle R_t + R_t^\dagger\rangle_t \Delta t + \Delta\beta_t, \tag{5.48}$$

where $\Delta x_t \equiv x_{t+\Delta t} - x_t$ and $\Delta\beta_t \equiv \beta_{t+\Delta t} - \beta_t$. The discretization of the loss function (5.45) is

$$\text{loss}(H, R, A, |\psi_0\rangle) = -A\sum_t \langle R_t + R_t^\dagger\rangle_t \Delta x_t + \frac{A^2}{2}\sum_t \langle R_t + R_t^\dagger\rangle_t^2 \Delta t. \tag{5.49}$$

Neglecting the constant $-{\Delta x_t}^2/2\Delta t$ and the multiplicative factor $\Delta t/2$, it is expressed as

$$\text{loss}(H, R, A, |\psi_0\rangle) = \sum_t \left(\frac{\Delta x_t}{\Delta t} - A\langle R_t + R_t^\dagger\rangle_t\right)^2. \tag{5.50}$$

Note that substituting $\Delta x_t/\Delta t$ by x_t, this is equal to Eq. (5.33).

5.5.5 Regularization

What should be the range of values of the learning parameters? If we had any intuition or knowledge about this question, we could use it to bias the learning. The way to bias or constrain learning is to introduce regularizers. This is equivalent to introducing a prior and doing maximum a posteriori instead of maximum likelihood.

Regularization of H

Consider the following discretized Scrödinger equation in the interaction picture

$$|\tilde{\psi}_{t+\Delta t}\rangle = \left[\mathbb{1} - \frac{\sigma^2}{2}R^\dagger(t)R(t)\Delta t + R(t)\Delta x_t\right]|\tilde{\psi}_t\rangle \tag{5.51}$$

where $R(t) \equiv e^{iHt}Re^{-iHt}$. If H is diagonal with eigenvalues ω_n, the matrix elements of R are $R_{ab}(t) = R_{ab}e^{i(\omega_a-\omega_b)t}$. Suppose we want to learn a single sequence, e.g., $x_t = \sin(\omega t)$ [i.e.,

$\Delta x_t/\Delta t \approx \omega\cos(\omega t)]$, such that the loss function is

$$\text{loss} = \sum_t \left(\omega\cos(\omega t) - A\langle R(t) + R^\dagger(t)\rangle_t\right)^2. \tag{5.52}$$

The loss function is minimised when $A\langle R(t)+R^\dagger(t)\rangle_t = \omega\cos(\omega t)$, i.e. when the expectation value $\langle R(t) + R^\dagger(t)\rangle_t$ oscillates with frequency ω. Since matrix elements of both $R(t)$ and $R^\dagger(t)R(t)$ oscillate with frequencies that are differences of eigenvalues of H, it is intuitive that the learned diagonal elements of H should be related to the frequency ω of the training data.

If we assume that H is related to the frequencies, it makes sense to limit it to the bandwidth of audio. *Nyquist's theorem* states that in order to correctly capture a discrete signal, the sampling rate must be at least double the highest frequency contained in the signal. Conversely, the highest frequency that can be captured at a given sampling rate is half the sampling frequency. This frequency is called the *Nyquist frequency*. If differences of eigenvalues of H give frequencies, the spectrum of H should be limited to $\pm s/4$, where s is the sampling rate. Thus if we set the standard deviation of the frequencies to be $\sigma_f = s/4$, and bearing in mind that $\omega = 2\pi f$, the regularization term in the loss should be

$$\mathcal{L}_H = \frac{1}{2\sigma_\omega^2}\sum_n \omega_n^2 = \frac{1}{8\pi^2\sigma_f^2}\sum_n \omega_n^2 = \frac{2}{\pi^2 s^2}\sum_n \omega_n^2, \tag{5.53}$$

which, up to a constant, corresponds to the logarithm of the Gaussian prior

$$p(\omega_1, ..., \omega_D) = \prod_{n=1}^{D} \sqrt{\frac{1}{2\pi\sigma_\omega^2}}\exp\left(-\frac{\omega_n^2}{2\sigma_\omega^2}\right). \tag{5.54}$$

Regularization of R

The scale of the signal is set by $A\langle R + R^\dagger\rangle_t$. If the typical scale of the matrix elements of R is r, its value should be determined by $\Delta x = Ar\Delta t$. If we set $A = 1$, we could introduce a Gaussian prior so that $\sigma_R = \Delta x/\Delta t$. The hyperparameter Δx can be inferred from the data. Then,

$$\mathcal{L}_R = \frac{1}{2\sigma_R^2}\sum_{ij} |r_{ij}|^2. \tag{5.55}$$

On the other hand, in general A is a learning variable and so it is not obvious what the regularizer of R should be.

5.5.6 Learnable parameters and hyperparameters

The model is specified by the learnable parameters A, H, R and $|\psi_0\rangle$ (or ρ_0) and the hyperparameters:

1. The bond dimension D, which reflects the complexity of the model.
2. Time discretization Δt. If we set it to be equal to the inverse of the sampling rate of the data, which corresponds to matching the time discretization of the data with the time discretization of the model, we can relate the eigenvalues H to the (angular) frequencies $\omega = 2\pi f$ of the data.
3. σ, which governs the strength of the term $R^\dagger R$.
4. The hyperparameters σ_ω and σ_R are regularisers for H and R.

5.5.7 Sampling

After training, we use the learned parameters H, R, A and $|\psi_0\rangle$ (or ρ_0) to generate samples using the discrete model

$$\begin{aligned} x_{t+1} &= x_t + \Delta x_t, \text{ where,} \\ \Delta x_t &= A\langle R_t + R_t^\dagger\rangle_t \Delta t + \sqrt{T}\Delta\beta_t. \end{aligned} \tag{5.56}$$

We introduce a temperature parameter T to tune the variance of the independent increments of the Brownian motion. In generative modeling, it is common to introduce a temperature parameter to optimize the quality of the sampling. See for example Ref. [25]. At $T = 0$ the generative process is deterministic. As the temperature is increased, the generative model gives rise to a variety of samples due to the randomness of the increments. At very high temperatures, the Gaussian noise dominates the generative process and the samples resemble the training data less and less.

5.6 Experiments

To test the capabilities of our model, we create synthetic datasets where we know the ground truth probability distributions. We can then readily check whether the learnt probability distribution matches the ground truth. We train on three different datasets: damped sines with random delays, *Gaussian processes* and *filtered Poisson processes*. The experiments consist on:

1. minimizing the negative log likelihood of the dataset with respect to the training variables H, R, A and $|\psi_0\rangle$ (or ρ_0),

2. generating new samples with the trained model,

3. comparing samples with training samples, either using the bare eye or computing different moments.

5.6.1 Damped sines

5.6.1.1 Experimental details

1. We consider the time derivative of the data as the homodyne current, i.e., $I_t = dx_t/dt$. The matrix R is complex and we set its diagonal elements to zero. Hence, we only keep oscillatory parts of $R(t)$, which we consider appropriate to model oscillatory data.

2. We learn the initial state $|\psi_0\rangle$ (or ρ_0). When using density matrices, we parameterize ρ_0 by $\rho_0 = \frac{W^\dagger W}{\mathrm{tr}[W^\dagger W]}$ to enforce normalization and real and positive eigenvalues. The matrix $W \in \mathbb{C}^{r\times D}$ defines the rank of the initial density matrix, with $r = 1$ corresponding to an initial pure state $|\psi_0\rangle$.

3. We use regularisers for the elements of H and R. These are set to $\sigma_\omega^2 = \frac{(16000\pi)^2}{400}$ and $\sigma_R^2 = 5$. They are included in the model as a refinement, but we do not experiment with them. They will become important when training on real data, which is more complex than the data considered here.

4. The hyperparameter dt remains fixed to $dt = 1/16000$. We experiment with different values of D and σ but only show results with the values that give the best results.

5. The batch size is 8.

5.6.1.2 Single frequency experiment

We start by modeling a dataset that consists of damped sines with random delays. Each signal has amplitude zero at the beginning, and the length of this "silence" period is random (see two samples in Fig. 5.1(a)). All signals have the same frequency $f = 261.6$Hz, the sampling rate is 16KHz and the length of each signal vector is 512 (which corresponds to 0.032 seconds). To generate the training set, we obtain the random delays by sampling from the distribution Gamma($\alpha = 2, \beta = 0.39$).

We start by considering the pure state model. The results are:

1. At $T = 0$, the sampling is deterministic given the learned initial state $|\psi_0\rangle$, as explained in Sec. 5.5.7. The zero temperature sample has the shape of a damped sine with a finite delay. This sample is shown in Fig. 5.1(b).

2. At finite temperature, we find that we can capture the delay degree of freedom, i.e. different samples have the form of a damped sine, with different delays. On the other hand, we find that the samples have the right form (i.e., the form of a damped sine) for the first 300 points only, having been trained on signals of length 512. In this sense, the outcome of this experiment is not very satisfactory. We experimented with different bond dimensions up to $D = 300$. We show two samples in Fig. 5.2(a).

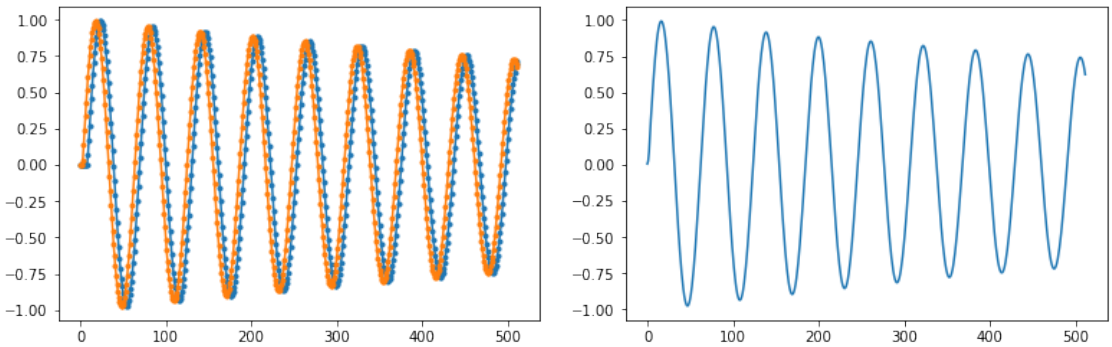


Figure 5.1: (a) Two signals of the dataset with different delays. The length of the data samples is 512 and the sampling frequency 16 kHz. (b) The $T = 0$ sample from the our pure state model, after training. It has the form of a damped sine with a finite delay.

We now consider the time evolution of a density matrix

$$\tilde{\rho}_{t+\Delta t} = \left[\mathbb{1} - \frac{\sigma^2}{2} R_t^\dagger R_t \Delta t + R_t \Delta x_t\right] \tilde{\rho}_t \left[\mathbb{1} - \frac{\sigma^2}{2} R_t^\dagger R_t \Delta t + R_t \Delta x_t\right]^\dagger . \tag{5.57}$$

In this case, sampling remains of good quality up to 512 samples, as can be seen in Fig. 5.2(b).

5.6.1.3 Two frequencies experiment

We want to see if we can generate samples of two different frequencies, after training on a dataset of damped sines with random delays and two different frequencies $f = 600, 800$Hz. The length of the training sequences is 100 samples.

We start with the pure state model. We train the model on a dataset that only contains two signals, shown in Fig. 5.3(a). After training, our model generates signals with different

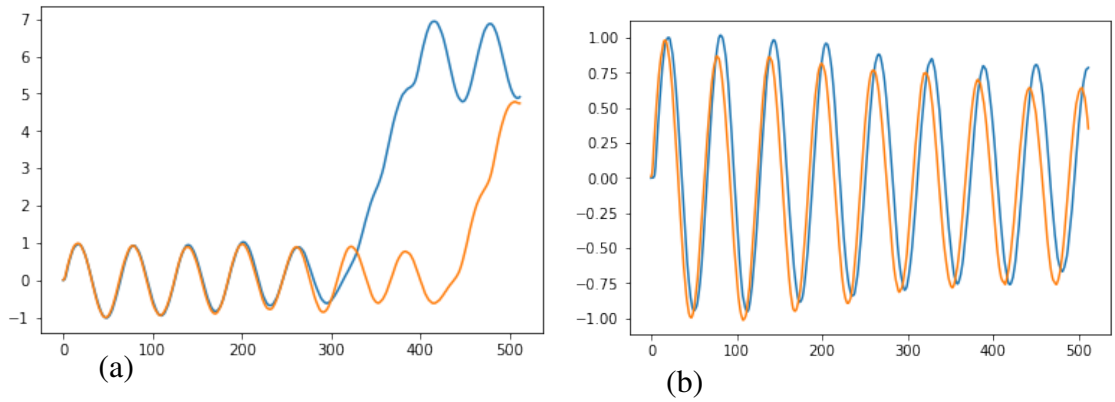


Figure 5.2: (a) Two samples at $T = 30$, $D = 100$ and $\sigma = 10^{-4}$ using the pure state model. The shape of a damped sine is well captured. On the other hand, we can only get proper samples of length 300 approximately. (b) Two samples at $T = 42$, $D = 100$ and $\sigma = 10^{-4}$, where we use a density matrix. Unlike in the pure state case, the samples look like damped sines, for the whole length of 512 samples.

frequencies that lie in between the two frequencies of the dataset, as shown in Fig. 5.3(b). The frequencies of the samples seem to be closer to $f = 800$Hz than to $f = 600$Hz.

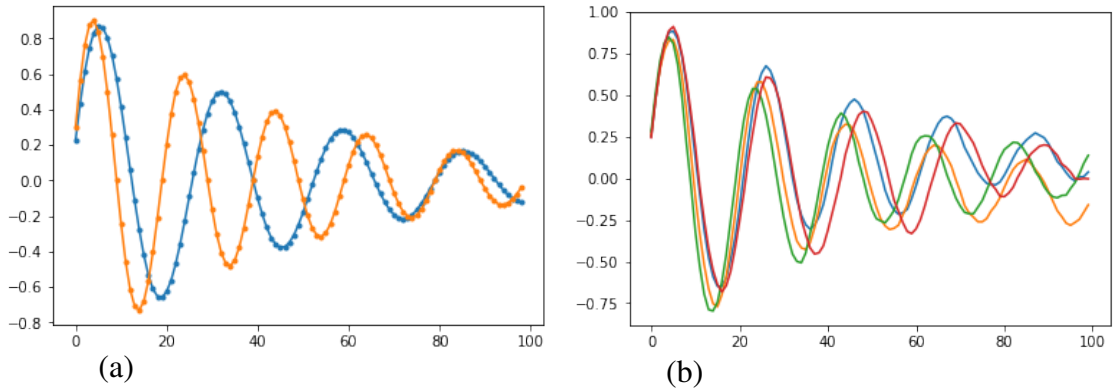


Figure 5.3: (a) Training set made of two signals. (b) Four samples at T=100 and $D = 50$ and $\sigma = 10^{-4}$, after having trained on the two signal dataset shown in (a).

We then move on to modeling a dataset of damped sines with random delays like we did in Sec. 5.6.1.2, but this time the training set will contain damped sines of two different frequencies as shown in Fig.5.4(a). We find that after training, samples are always quite close to the higher frequency. Different generated samples have different shapes, but all look like damped sines. The model learns the manifold of damped sines fairly well, but it fails to capture the two frequencies degree of freedom of the dataset, in that there are no samples with frequencies close to $f = 600$Hz. We show the result in Fig. 5.4(b).

We experiment considering the time evolution of a density matrix but the performance of the model does not improve compared to the pure state case, i.e. it fails to capture the

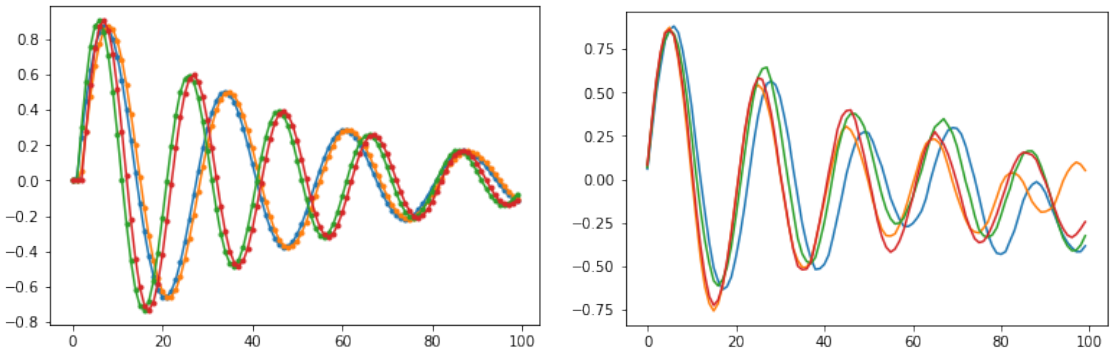


Figure 5.4: (a) Four signals of the dataset with different delays and frequencies $f = 600, 800$ Hz. The length of the data samples is 100 and the sampling frequency 16 kHz. (b) Four samples at $T = 120, D = 100$ and $\sigma = 10^{-4}$.

two frequencies degree of freedom of the dataset.

5.6.2 Gaussian processes

In the previous section, we tested the ability of our model to learn damped sines. On the other hand, real life sound is a lot more complex than sine waves. For example, real sound is made of several harmonics (unlike a sine wave). To test the capabilities of our model on more realistic data, we move on to training on *Gaussian processes*.

A stochastic function $x(t)$ is a Gaussian process (GP) if any finite collection of random variables $x(t_1), ..., x(t_n)$ have a multidimensional Gaussian distribution [123]. A GP is defined in terms of its *mean* $\mathrm{m}(t)$ and its *covariance function* (or *kernel*) $\mathrm{C}(t, t')$, defined as

$$\mathrm{m}(t) = \mathbb{E}[x(t)], \tag{5.58}$$

$$\mathrm{C}(t, t') = \mathbb{E}\left[(x(t) - \mathrm{m}(t))\,(x(t') - \mathrm{m}(t'))\right]. \tag{5.59}$$

A Gaussian process is *stationary* if the mean is time independent and the covariance function only depends on time differences

$$\mathrm{C}(t, t') = \mathrm{C}(t - t'). \tag{5.60}$$

We use the notation $\mathrm{C}(\tau)$ (where $\tau \equiv t - t'$) when considering stationary processes. The Wiener-Khintchine theorem relates the stationary kernel to a corresponding spectral function

$$S(\omega) = \int_{-\infty}^{\infty} d\tau \, \mathrm{C}(\tau) e^{-i\omega\tau}, \tag{5.61}$$

$$\mathrm{C}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \, S(\omega) e^{i\omega\tau}. \tag{5.62}$$

A specific kind of stationary GPs that have been used to reflect the complex harmonic structure of musical notes are *Matérn spectral mixtures*. They have been used for different sound related machine learning tasks [124]. Consider the kernels

$$\mathrm{C}_{1/2}(\tau) = \sigma^2 e^{-\lambda\tau}, \tag{5.63}$$

$$\mathrm{C}_{\cos}(\tau) = \cos(\omega_0 \tau). \tag{5.64}$$

The corresponding spectral densities are

$$S_{1/2}(\omega) = \frac{2\sigma^2\lambda}{\lambda^2 + \omega^2}, \tag{5.65}$$

$$S_{\cos}(\omega) = \pi \left[\delta(\omega - \omega_0) + \delta(\omega + \omega_0)\right]. \tag{5.66}$$

The spectral density of the product of the two kernels $C(\tau) = C_{1/2}(\tau)C_{\cos}(\tau)$ takes the form of a pair of Lorentzians centered at $\pm\omega_0$

$$S(\omega; \boldsymbol{\theta}) = 2\pi\sigma^2\lambda \left[\frac{1}{\lambda^2 + (\omega - \omega_0)^2} + \frac{1}{\lambda^2 + (\omega + \omega_0)^2}\right], \tag{5.67}$$

where $\boldsymbol{\theta} = (\sigma, \lambda, \omega_0)$. The general form of *Matérn spectral mixtures* is a sum over different pairs of Lorentzians

$$S_{\mathrm{SMS}}(\omega; \boldsymbol{\Theta}) = \sum_{j=1}^{N} S(\omega; \boldsymbol{\theta}_j), \tag{5.68}$$

where $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_j\}$. The corresponding covariance is

$$\mathrm{C}_{\mathrm{MSM}}(\tau; \boldsymbol{\Theta}) = \sum_{j=1}^{N} \sigma_j^2 e^{-\lambda_j\tau} \cos(\omega_j \tau). \tag{5.69}$$

5.6.2.1 Experimental details

1. We consider the data to be the homodyne current, i.e. $I_t = x_t$. This is because as shown in Fig. 5.5, on this dataset the increments of the signal are a lot more spiky than the signals themselves, which makes learning difficult in the continuous formulation of the problem.

2. We learn the initial state $|\psi_0\rangle$.

3. The hyperparameter dt is set to $dt = 0.001$ and $\sigma = 1$. We experiment with different values of D and σ but only show results with the values that give the best results.

4. We do not use regularisers.

5. The batch size is 8.

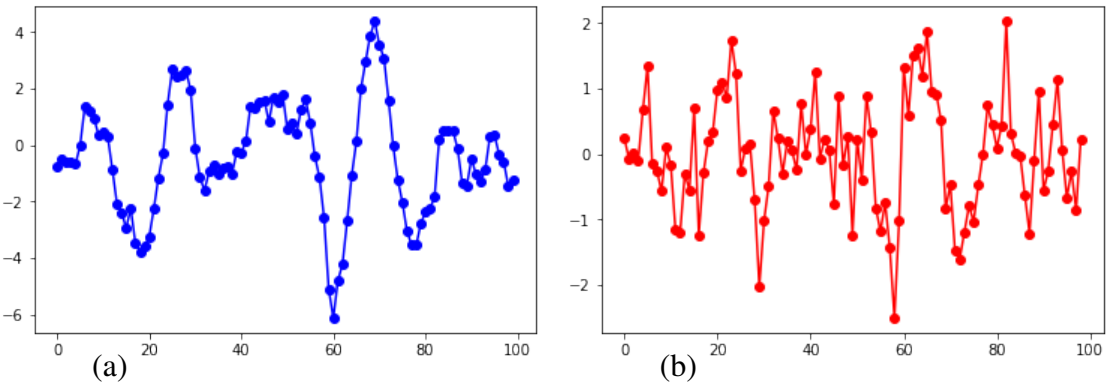


Figure 5.5: (a) A sample x_t of a Gaussian process with *Matérn spectral mixture* spectral function defined by $(\sigma, \lambda, \omega_0) = (2, 50, 300)$. (b) The increments $\Delta x_t = x_{t+dt} - x_t$ of the sample shown in (a).

We create a dataset of samples of a stationary Gaussian process of choice. We generate the data using a discrete stochastic equation (see Appendix. B.1.1) instead of sampling from a multivariate Gaussian distribution. We train on two different *Matérn spectral mixture* processes. In the first, the spectral function consists of a single pair of Lorentzians centered at $\omega_0 = \pm 300$ and $(\sigma, \lambda) = (2, 50)$. In the second, we consider a mixture of three frequencies. The mixture is defined by the parameters $(\sigma_i, \lambda_i) = (2, 50)$ for $i = 1, 2, 3$ and $(\omega_1, \omega_2, \omega_3) = (300, 500, 700)$. We show two samples from each of the two datasets in Fig. 5.6.

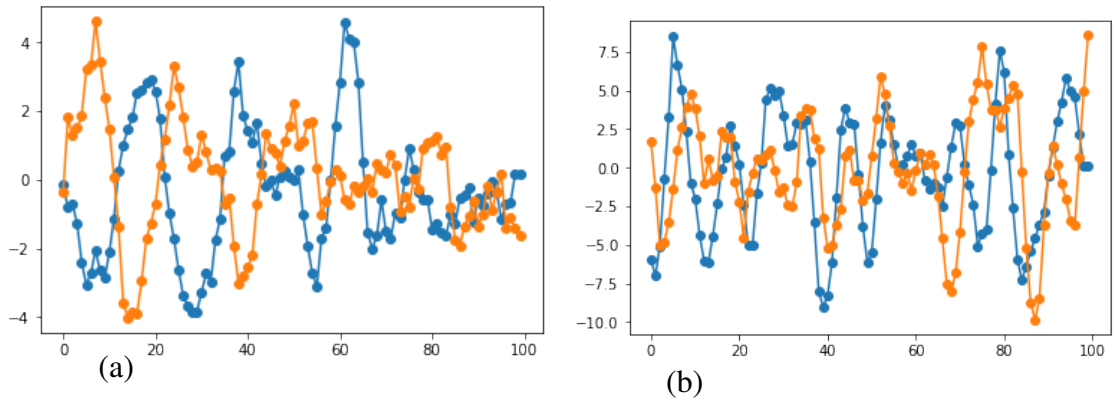


Figure 5.6: Samples from the two dataset we train on. Two Gaussian processes with *Matérn spectral mixture* spectral function defined by (a) $(\sigma, \lambda, \omega_0) = (2, 50, 300)$ and (b) $(\sigma_i, \lambda_i) = (2, 50)$ for $i = 1, 2, 3$ and $(\omega_1, \omega_2, \omega_3) = (300, 500, 700)$.

5.6.2.2 Results

Due to the higher complexity of the data compared to the damped sines in Sec. 5.6.1, instead of just looking at plots of samples, we judge whether the model is successful at learning the above process by 1) calculating the experimental covariance from N samples

$$C_{\text{exp}}(t, t') \equiv \frac{1}{N} \sum_{i=1}^{N} x_i(t) x_i(t'), \tag{5.70}$$

and comparing it with the exact covariance and 2) checking that the experimental covariance is stationary, i.e. $C_{\text{exp}}(t, t') = C_{\text{exp}}(\tau)$. We find that the model is successful at learning this process and we show the results on Fig. 5.7. On the other hand, as explained in Sec. 5.5.7, sampling depends on temperature. The experimental covariance only matches the exact covariance at a given temperature. As one departs from this temperature, the two covariances start to differ. Similarly, the samples are stationary only in a small range of temperatures around this temperature. Furthermore, we find that the experimental covariance becomes stationary only after a few steps, not from the beginning.

When trained on a mixture of three frequencies, the model succeeds at reproducing stationary samples (after a given time) with the right covariance function. The results are shown in Fig. (5.8).

5.6.3 Poisson processes

A feature of stationary Gaussian processes is that because the covariance function is symmetric $\mathrm{C}(t, t') = \mathrm{C}(t', t)$ and all diagonal elements are equal to $\mathrm{C}(t, t)$, the probability density of a given sample $x(t)$ is the same as the probability density of the time-inverted

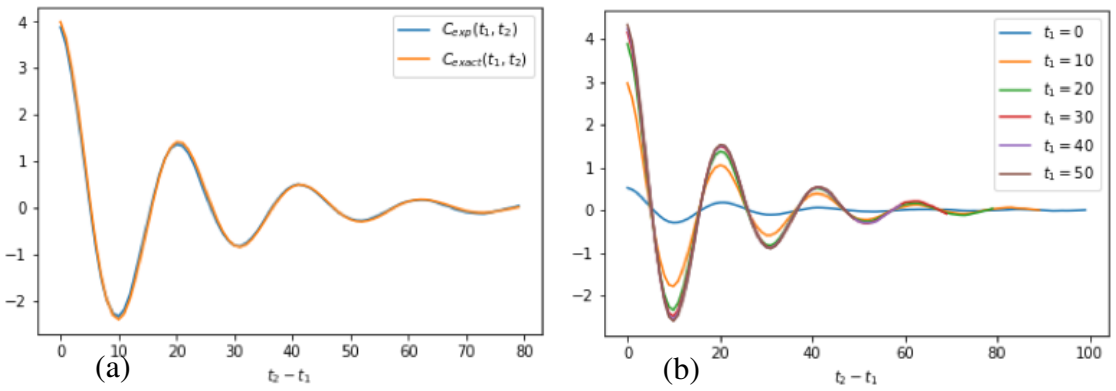


Figure 5.7: The time indicated in the horizontal axis is an integer index that specifies the time step of the discrete covariance. (a) Perfect match of experimental (blue) and exact (yellow) covariances. The exact covariance has parameters $(\sigma', \lambda, \omega) = (2, 50, 300)$. The experimental covariance is calculated using 40000 samples at $T = 0.00051$. The hyperparameters used are $D = 50$, $dt = 0.001$ and $\sigma = 1$. (b) Experimental covariance $C_{exp}(t_1, t_2)$ for different initial times, showing stationarity. It reaches stationarity at $t_1 \approx 20$.

sample. We refer to this symmetry as *time-reversal symmetry* (TRS).

Many real life sounds are not time-reversal symmetric. As an example, consider bird chirping. If one records bird chirping and run the record backwards, it will not sound like bird chirping. Therefore time-reversal symmetric models like Gaussian models are not suitable to model this kind of sound.

Our cMPS based model is not constrained by time-reversal symmetry, as multivariate Gaussian probability distributions are. We can see this by looking at the discretized time evolution of the unnormalized state. The fact that the one-step time evolution operator does not commute with itself at different times, implies the absence of the TRS constraint.

One can check whether a probability distribution is time-reversal symmetric, from certain correlation functions. Consider the two correlators

$$\begin{aligned}
\mathbb{E}\left[x^3(t_i)x(t_j)\right] &= \int dx(t_1)...dx(t_N)\; x^3(t_i)x(t_j)\; p\left(x(t_1), ..., x(t_i), ..., x(t_j), ..., ..., x(t_N)\right), \\
\mathbb{E}\left[x(t_i)x^3(t_j)\right] &= \int dx(t_1)...dx(t_N)\; x(t_i)x^3(t_j)\; p\left(x(t_1), ..., x(t_i), ..., x(t_j), ..., ..., x(t_N)\right).
\end{aligned} \tag{5.71}$$

If these two quantities are different, the probability is not invariant under the swap of values of two arguments which implies that it is not TRS.

We test the ability of our model to learn non-TRS processes, by training it on *filtered Poisson processes*. This process is defined as

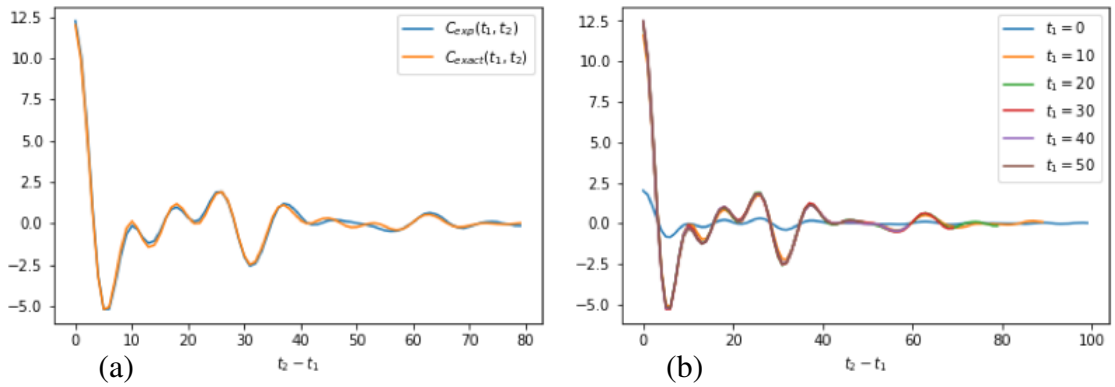


Figure 5.8: The time indicated in the horizontal axis is an integer index that specifies the time step of the discrete covariance. (a) Perfect match of experimental (blue) and exact (yellow) covariances. The exact covariance has parameters $(\sigma_i', \lambda_i) = (2, 50)$ for $i = 1, 2, 3$ and $(\omega_1, \omega_2, \omega_3) = (300, 500, 700)$. The experimental covariance is calculated using 40000 samples at $T = 0.002$. The hyperparameters used are $D = 100$ and $dt = 0.001$. (b) Experimental covariance $C_{exp}(t_1, t_2)$ for different initial times, showing stationarity. Before $t_1 = 10$ the experimental covariance is non-stationary.

$$X(t) = \sum_k A_k \varphi(t - t_k), \text{ where}$$
$$\varphi(t - t_k) = \theta(t - t_k) e^{-(t-t_k)/\tau} \sin[\omega(t - t_k)]. \tag{5.72}$$

A *filtered Poisson process* (FPP) $X(t)$ consists of a superposition of uncorrelated pulses $\varphi(t - t_k)$, arriving at random times with a Poisson distriution. The overall amplitude A_k is random: at each time, A_k can independently take the values $\pm A$, with equal probabilities. In this process, the correlators defined in Eq. (5.71) take the form (see the derivation in Appendix B.2)

$$\mathbb{E}\left(X^3(t_1)X(t_2)\right) = \lambda I_{3,1}^{-\infty,t_1} + 3\lambda^2 I_{1,1}^{-\infty,t_1} I_{2,0}^{-\infty,t_1},$$
$$\mathbb{E}\left(X(t_1)X^3(t_2)\right) = \lambda I_{1,3}^{-\infty,t_1} + 3\lambda^2 I_{1,1}^{-\infty,t_1} \left(I_{0,2}^{-\infty,t_1} + I_{0,2}^{t_1,t_2}\right),$$
$$I_{n,m}^{t,t'} = \int_t^{t'} d\alpha\, \varphi^n(t_1 - \alpha)\varphi^m(t_2 - \alpha). \tag{5.73}$$

The two correlators are different due to the absence of TRS. We take the initial time $t = -\infty$ because we are interested in the steady state correlators.

The experimental details are the same as in 5.6.2.1. The dataset contains samples of the FPP defined in Eq. (5.72) with parameters $A = 1$ (i.e., A_k can take values ± 1 at time

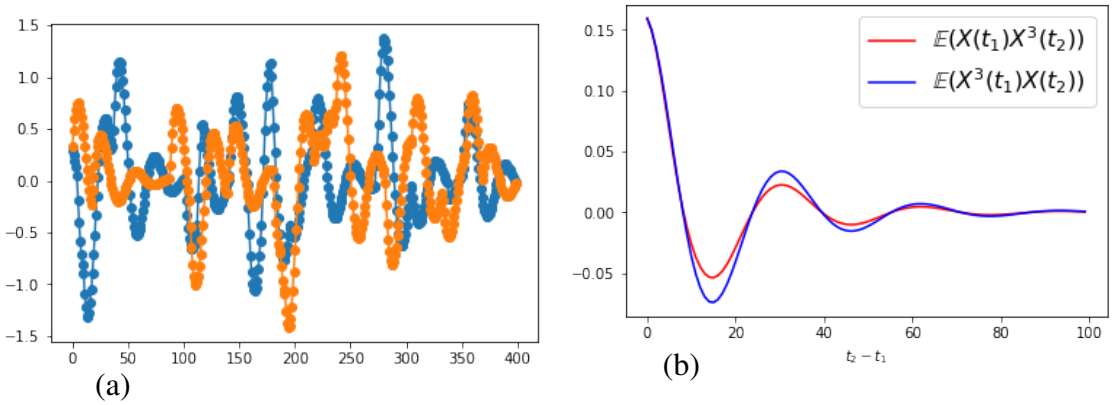


Figure 5.9: (a) Training samples generated according to the FPP defined in Eq. (5.72). (b) Exact correlators of FPP defined in Eqs. (5.73) of the Poisson process defined in Eq. (5.72). The intensity of the Poisson process is $\lambda = 4$. The amplitude A_k can take values ± 1. The pulse decay time is $\tau = 0.2$ and the angular frequency $\omega = 20$. The time indicated in the horizontal axis is an integer index that specifies the step of the discrete correlator.

t_k), $\tau = 0.2$, $\omega = 20$. In order to create steady state signals, we produce signals of length 500, and pick the last 400 points of each signal. This corresponds to signals that have been running for a time 5τ, by when signals are approximately stationary, because the process has a memory time of order τ due to the exponential decay of the pulses $\varphi(t - t_k)$. The Poisson intensity parameter is $\lambda = 4$. We show two steady state training samples in Fig. 5.9(a).

5.6.3.1 Results

By tuning the temperature, we can match the experimental correlators to either of the two exact correlators. On the other hand, both experimental correlators we obtain are equal.

5.7 Conclusions

We introduce a quantum-inspired generative model for raw audio. Our model takes the form of a stochastic Schrödinger equation describing the continuous time measurement of a quantum system. This constitutes a deep autoregressive architecture in which the system's state is a latent representation of past observations.

We rephrase the model in the language of stochastic differential equations. We derive an expression to calculate the two-time characteristic function of a filtered Poisson process. We test our model on three different synthetic datasets:

1. Damped sines:

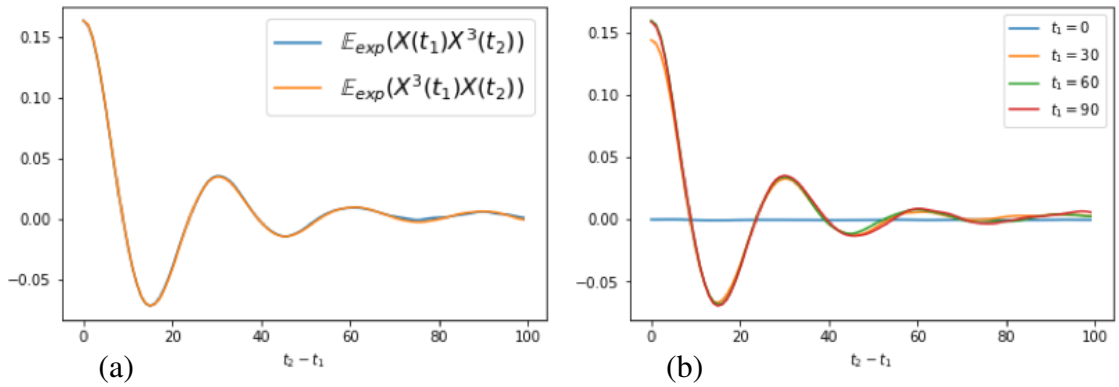


Figure 5.10: The time indicated in the horizontal axis is an integer index that specifies the time step of the discrete correlator. (a) Experimental correlators $\mathbb{E}_{\text{exp}}\left(X(0)X(t)^3\right)$ and $\mathbb{E}_{\text{exp}}\left(X^3(0)X(t)\right)$. At $T = 0.000012$, they roughly match the exact correlator $\mathbb{E}\left(X^3(t_1)X(t_2)\right)$ shown in Fig. 5.9(b). The temperature can be tuned to approximately match $\mathbb{E}\left(X(t_1)X^3(t_2)\right)$ instead. But at a given temperature they are both equal, unlike the exact correlators. The hyperparameters used are $dt = 0.01$, $D = 100$ and $\sigma = 1$. The experimental correlators are calculated by averaging over 40000 samples. (b) The experimental correlator $\mathbb{E}_{\text{exp}}\left(X(t_1)X(t_2)^3\right)$, for different values of t_1. It becomes stationary after $t_1 = 30$, approximately.

We start by training on damped sines of a single frequency with random delays. We are able to learn the manifold of damped sines of the right frequency with different delays. On the other hand, in order to get good quality samples with the length of training samples, the pure state model is not successful but we need to consider the evolution of a density matrix, which is successful at learning and generating good samples of the full length.

We then include two different frequencies of the of damped sines with random delays. Here, the pure state model captures the manifold of damped sines with random delays. On the other hand, the generated samples have frequencies similar to the higher frequency of the training set, failing to capture the two frequency degree of freedom of the training data. When we consider the evolution of a density matrix the performance does not improve.

2. Gaussian processes:

We train on two different datasets made of samples of two different stationary Gaussian processes of the *Matérn spectral mixture* kind. The pure state model is able to learn the right manifold. We test this by computing the experimental covariance function and comparing with the exact result. We find that the SDE model where we consider the time derivative of the data to be the homodyne current, is not suitable for this dataset because the time derivative of the signals are more discontinuous

than the signals themselves.

3. Filtered Poisson processes:

   We train on a dataset made of samples of a *filtered Poisson process*. We capture the right correlators $\mathbb{E}\left(X^3(t_1)X(t_2)\right)$ and $\mathbb{E}\left(X(t_1)X^3(t_2)\right)$ by tuning the sampling temperature T, but we fail to discern between the two, i.e., both experimental correlators are equal. By tuning temperature, we can make them roughly match the exact expression of one or the other, but not both.

It remains to do a proper hyperparameter tuning considering all the hyperparameters, to see if the performance of the model can be improved. Moreover, and most importantly, the model needs to be tested on real data: how expressive is the model and how is this related to the bond dimension? How is the quantum entanglement of the model related to the the structure of correlations in the generated samples?

This work opens a new avenue to use matrix product states to model continuous data and we hope that it will set the beginning of the exploration of cMPS for machine learning.

6

Conclusions and outlook

In this thesis we have presented research done on three different topics: vortices in superfluids, quantum many-body optimization and machine learning. All projects share the feature of being related to quantum many-body physics. There has been an evolution in the way our work is related to many-body physics: we started by studying a many-body quantum system, we then moved to developing a technique to solve quantum many-body problems and finally we used ideas from quantum physics to do machine learning.

In Chapter 2 we studied the geometry of infinite lattices of vortex molecules, in the ground state of Rabi-coupled condensates. We found qualitative agreement with previous results in the bulk of harmonic traps. More interestingly, we came up with an effective theory in terms of point charges and derived an expression for the interaction energy density in terms of an elliptic function $\mathcal{Q}(z_{ij};\omega_1,\omega_2)$, in the limit of point vortices and very small molecules when the net quadrupole moment is zero. We used this expression to find the ground state configuration of an infinite system of molecules.

In Chapter 3 we studied the ground state of a two component spinor BEC. In the polar phase, where the spin density is zero, the emergent *spin-gauge rotation symmetry* allows for the presence of half-quantum vortices. We found HQVs in the ground state, in certain region of the parameter space covered. We provided analytical arguments to explain that in the polar phase, a pair of half-vortices is energetically favorable compared to a single integer vortex, in the logarithmic approximation.

In Chapter 4 we proposed a new MPS based method to find ground states of one-dimensional quantum lattices. We identified a connection between matrix product states in the left canonical form and the Stiefel manifold. This allows to constrain the optimisation to this subspace of the otherwise larger MPS manifold. We found that our method suffers from two undesirable features. First, in order to get beyond single precision accuracy, the size of the unit cell needs to be increased and so we could not get the acuraccy that Zauner-Stauber *et al.* [104] achieve using a translationally invariant MPS (i.e., single-site unit cell). Second, because of the presence of the power method in the variational energy expression, it is possible for the convergence process to get stuck in regions of the Stiefel

manifold where the modulus of the second largest eigenvalue of the transfer matrix is very close to one. It remains to benchmark this method with others. There is a chance that even with the required large unit cell, our algorithm is still faster than other existing methods like DMRG or the algorithm introduced by Zauner-Stauber *et al*. in Ref. [104].

In Chapter 5 we introduced a generative model for raw audio, based on continuous matrix product states. Our model takes the form of a *stochastic Schrödinger equation* describing the continuous time measurement of a quantum system. Moreover, our model constitutes a deep *Recurrent Neural Network* (RRN). This kind of neural network is well suited for the modeling of sequential data, like raw audio. There, the *hidden state* which is a function of all previous observations, plays the role of a *memory* that captures correlations across the sequence. In our model, this is incarnated by a wave function or a density matrix, which depends on all past measurements of the quantum system. The model is more expressive when considering the time evolution of a mixed state than when considering a pure state. On the other hand, it becomes computationally more expensive due to the larger number of training variables.

Specifically, we considered the continuous time measurement of the homodyne current at the output of an open quantum system. We considered the homodyne current to be either the data or the time derivative of the data we wanted to model. When we considered the second, we could rephrase the model in the language stochastic differential equations, giving a continuous formulation of the problem. Note that raw audio data in digital form corresponds to discretised continuous waves.

We tested our model on three different synthetic datasets. The first dataset consisted of damped sines with random delays. Here we found that the model performed well only when we considered the time evolution of a density matrix. The pure state model failed to give good quality samples with a length equal to the length of the training samples. Different generated samples had the shape of a damped sine with different delays, like the training set.

When we included damped sines with two different frequencies in the training dataset, the model did not perform as well. It captured well the manifold of damped sines in that, after training, generated samples had the shape of a damped sine. On the other hand, the model (both pure and mixed state) failed to capture the two-frequency degree of freedom of the data. All the samples had a frequency that was very similar to the higher frequency of the sine-waves we trained on.

We then trained on a dataset composed of samples of a *Matérn spectral mixture* stationary Gaussian process. Here we judged the success of the model by computing the experimental covariance of the generated samples and comparing with the exact covariance of the training set. We found that by tuning the sampling temperature, we could very

accurately recover the exact covariance from generated samples.

Finally, we trained on a dataset made of samples from a *filtered Poisson process* (FPP). We chose this dataset because its corresponding probability density is not constrained by *time-reversal symmetry* (TRS). We defined time reversal symmetry as a property of a probability distribution, where the probability of a given sample is equal to the probability of the time-inverted sample. Stationary Gaussian models are time-reversal symmetric. This constraint is not well suited to many real life sounds, like bird chirping for example. The initial attack of the chirp with the subsequent decay makes this sound inherently non-TRS. Our model is not constrained by time-reversal symmetry. For non-TRS probability distributions, the correlators $\mathbb{E}\left(X^3(t_1)X(t_2)\right)$ and $\mathbb{E}\left(X(t_1)X^3(t_2)\right)$ are different. We judged the capability of our model for learning FPPs by computing these two correlators from the generated samples and comparing with the exact result. The model succeeded at recovering either of these correlators from generated samples, by tuning the sampling temperature. On the other hand, it failed to discern between the two. As a function of temperature, we could roughly match one or the other, but not both.

It remains to do a proper hyperparameter tuning considering all the hyperparameters, to see if the performance of the model can be improved. Moreover, and most importantly, the model needs to be tested on real data to see if it is useful for real world applications.

In the last part of this dissertation, we have brought ideas across from many-body quantum mechanics to the field of machine learning. Specifically, we have borrowed a numerical tool (cMPS), that is used to deal with the gigantic Hilbert space of quantum many-body systems, to fight a similar problem that appears in the context of raw audio modeling: the *curse of dimensionality*. We have used our quantum-inspired model to learn different kinds of data, showing the potential of this approach to be applied to a broad range of real-world problems. We hope that this work will motivate more work along this line and we are very excited to see what physics can do to push the boundaries of artificial intelligence.

A

Vortex related calculations

A.1 Derivation of boundary conditions

Let us parametrize the wave functions as $\psi = \sqrt{\rho}e^{i\theta}$. We define the lattice vectors $\mathbf{R}_x = L_x\hat{x}$ and $\mathbf{R}_y = L_y\hat{y}$. L_x and L_y are the dimensions of the computational unit cell.

We first require the density ρ_σ in each component to be periodic. This gives us a condition on the amplitudes of the wave-functions:

$$\sqrt{\rho_\sigma(\mathbf{r}+\mathbf{R}_i)} = \sqrt{\rho_\sigma(\mathbf{r})},\ \sigma = a,b,\ i = x,y. \tag{A.1}$$

We next require the superfluid velocity $\mathbf{v}_\sigma = \nabla\theta_\sigma - \mathbf{A}$ on each of the components to be periodic. Note that $\mathbf{v}_\sigma$ is the σ component superfluid velocity only when $\Omega_\mathrm{R} = 0$. On the other hand, it is still a gauge invariant quantity when $\Omega_\mathrm{R} \neq 0$ and hence its periodicity should be required:

$$\mathbf{v}_\sigma(\mathbf{r}+\mathbf{R}_i) = \mathbf{v}_\sigma(\mathbf{r}),\ \sigma = a,b,\ i = x,y. \tag{A.2}$$

This does still not give an explicit condition on the phase (it is a set of equations involving gradients of phases). In order to find it, we need to integrate the above equations. By doing so, we arrive to four equations:

$$\theta_a(\mathbf{r}+\mathbf{R}_x) = \theta_a(\mathbf{r}) + \Omega L_x y + \alpha_x, \tag{A.3}$$
$$\theta_a(\mathbf{r}+\mathbf{R}_y) = \theta_a(\mathbf{r}) - \Omega L_y x + \alpha_y, \tag{A.4}$$
$$\theta_b(\mathbf{r}+\mathbf{R}_x) = \theta_b(\mathbf{r}) + \Omega L_x y + \beta_x, \tag{A.5}$$
$$\theta_b(\mathbf{r}+\mathbf{R}_y) = \theta_b(\mathbf{r}) - \Omega L_y x + \beta_y. \tag{A.6}$$

Hence we are left with four integration constants.

Finally, we require periodicity of spin. Let's define the pseudo-spin density $\mathbf{S} = \Psi^\dagger \boldsymbol{\sigma} \Psi/2$, where $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ are the Pauli spin matrices. We first require periodicity of $S_x(\mathbf{r})$:

$$S_x(\mathbf{r} + \mathbf{R}_i) = S_x(\mathbf{r}), \ i = x, y. \tag{A.7}$$

These yields the conditions

$$\beta_x = \alpha_x + 2\pi n, \tag{A.8}$$

$$\beta_y = \alpha_y + 2\pi n, \ n \text{ integer}. \tag{A.9}$$

This condition also ensures the periodicity of $S_y(\mathbf{r})$ as well. The periodicity of $S_z(\mathbf{r})$ is ensured by the periodicity of the densities. Thus we are left with two integration constants α_x, α_y. The only effect of varying these two constants is to shift the wave-functions in the unit cell. These degrees of freedom are expected, since one reproduces the same infinite lattice upon copying unit cells, no matter where vortices sit in the computational unit cell. We choose $\alpha_x = \alpha_y = 0$. Note that we have still not said anything about the periodicity of true spin. The key point is that, because $\{\sigma_0, \sigma_x, \sigma_y, \sigma_z\}$ form a basis in the space of 2×2 matrices, periodicity of $\mathbf{S}(\mathbf{r})$ automatically ensures periodicity of spin. Hence, in order to fully fix the boundary condition for the wave-functions, it is necessary and sufficient to require periodicity of $\rho_\sigma, \mathbf{v}_\sigma$ and $\mathbf{S}$. The boundary condition is:

$$\begin{aligned} \psi_\sigma(\mathbf{r} + \mathbf{R}_x) &= e^{i\Omega L_x y} \psi_\sigma(\mathbf{r}) \\ \psi_\sigma(\mathbf{r} + \mathbf{R}_y) &= e^{-i\Omega L_y x} \psi_\sigma(\mathbf{r}), \ \sigma = a, b. \end{aligned} \tag{A.10}$$

A.2 Ω: allowed values and relation to the number of vortices

Let's prove that in order to avoid having a contradicting theory, the angular velocity can only take a discrete set of values. Using Eq. (A.10):

$$\begin{aligned}\psi(x+L_x,y+L_y) &= e^{-i\Omega L_y(x+L_x)}\psi(x+L_x,y)\\ &= e^{-i\Omega L_y(x+L_x)}e^{i\Omega L_x y}\psi(x,y)\\ \psi(x+L_x,y+L_y) &= e^{-i\Omega L_x(y+L_y)}\psi(x,y+L_y)\\ &= e^{i\Omega L_x(y+L_y)}e^{-i\Omega L_y x}\psi(x,y)\end{aligned} \tag{A.11}$$

$$\psi(x+L_x,y+L_y)=\psi(x+L_x,y+L_y)\Leftrightarrow\Omega=\frac{\pi n}{A}, \tag{A.12}$$

where n is an integer and $A = L_x L_y$.

In his seminal work, Feynman explains that the lowest energy state for an irrotational fluid with a given angular momentum is a vortex lattice, with a 2π winding of the phase around each vortex [4]. Because the superfluid velocity is $\mathbf{v} = \nabla\theta$, the superfluid cannot rotate as a rigid body. On the other hand, the vortex lattice (the set of vortex cores) can only rotate as a rigid body in equilibrium [32]. Hence, on average the region of the superfluid that is packed with vortices, rotates as a rigid body. This allows to estimate a relation between the angular velocity of the trap Ω and the number of vortices N_v, in the ground state. Let D be a region of area A packed with N_v vortices and $\mathbf{v} = \Omega r\hat{\varphi}$ (rigid solid rotation). Let ∂D be its boundary. If we calculate the circulation of the velocity:

$$\left.\begin{aligned}\Gamma_{\partial D}=\oint_{\partial D}\mathbf{v}\cdot d\mathbf{l}=2\Omega A\\ \Gamma_{\partial D}=2\pi N_v\end{aligned}\right\}\Rightarrow\Omega=\frac{\pi N_v}{A}.$$

From this result, we infer the meaning of n in Eq. (A.12): $n = N_v$. Even though this estimate is based on heuristic arguments, it is verified in the numerics, i.e. the number of vortices found in the ground state is n.

A.3 Numerical integration

We have an integral in two dimensions over the area A

$$E=\int_A d\mathbf{r}\mathcal{E}(\mathbf{r}). \tag{A.13}$$

To calculate the integral numerically, we use the *Riemann sum* method. We have a 2D grid of points (x_n, y_m), where $x_n = na_x$, $n, m \in [0, N-1]$, $y_n = na_y$. The discretization lattice constant is $a_x = L_x/N$ and $a_y = L_y/N$. Using this method, we approximate the integral as

the sum of the volumes of the parallelepipeds:

$$V_{nm} = a_x a_y \mathcal{E}(x_n, y_m), \tag{A.14}$$

and

$$E(\text{estim.}) = a_x a_y \sum_{n,m} \mathcal{E}(x_n, y_m). \tag{A.15}$$

In an analogous way to what is done in [125], we find that to leading order, the discretization error is

$$\Delta E = E(\text{estim.}) - E = -\frac{A}{2}(a_x \langle \partial_x \mathcal{E} \rangle + a_y \left\langle \partial_y \mathcal{E} \right\rangle), \tag{A.16}$$

where $\langle \partial_x \mathcal{E} \rangle = \frac{1}{N^2} \sum_{nm} \partial_x \mathcal{E}(x_n + a_x/2, y_m + a_y/2)$.

Let's now write the discrete version of Eq. (2.5), with $\omega_{\text{eff}} = 0$. It is all trivial to write except the kinetic energy term, so let's focus on that first. We introduce the vector potential by making the Peierls substitution:

$$\begin{aligned}
&\psi^*(x)(p_x - A_x)^2\psi(x) = \\
&\frac{1}{a^2}(2|\psi(x)|^2 - e^{iA_x a}\psi^*(x+a)\psi(x) - e^{-iA_x a}\psi^*(x)\psi(x+a)) \\
&+ \mathcal{O}(a^3).
\end{aligned} \tag{A.17}$$

Now, by defining the discrete wave function $\varphi(n, m) = \psi(x, y)\sqrt{a_x a_y}$:

$$\begin{aligned}
&E_d(\mathcal{R}, \{\varphi_\sigma(n,m)\}) = \\
&\sum_\sigma \sum_{n,m} \Big[\frac{1}{2a_x(\mathcal{R})^2} \left|\varphi_\sigma(n+1,m)e^{-iA_x a_x(\mathcal{R})} - \varphi_\sigma(n,m)\right|^2 \\
&+ \frac{1}{2a_y(\mathcal{R})^2} \left|\varphi_\sigma(n,m+1)e^{-iA_y a_y(\mathcal{R})} - \varphi_\sigma(n,m)\right|^2 \\
&- \mu_\sigma |\varphi_\sigma(n,m)|^2 \Big] \\
&+ \sum_{\sigma_1,\sigma_2} \frac{g_{\sigma_1\sigma_2}}{2} \sum_{n,m} \frac{1}{a_x(\mathcal{R})a_y(\mathcal{R})} |\varphi_{\sigma_1}(n,m)|^2 |\varphi_{\sigma_2}(n,m)|^2 \\
&- \Omega_{\text{R}} \sum_{n,m} \left[\varphi_a^*(n,m)\varphi_b(n,m) + \varphi_b^*(n,m)\varphi_a(n,m)\right].
\end{aligned} \tag{A.18}$$

Note that we have added two Lagrange multipliers $\{\mu_a, \mu_b\}$ to constrain the norm of the

wave function (and therefore the number of particles) in the minimization.

Here something important pointed out by Mingarelli *et al* is that in order to allow the vortex lattice configuration access any lattice geometry, one needs to minimize with respect to the aspect ratio of the unit cell $\mathcal{R} = a_x/a_y$. It is important to parametrize the discretization lattice constants in such a way that the area of the computational unit cell does not depend on it. Such a parametrization is:

$$a_x(\mathcal{R}) = \sqrt{\frac{A}{N^2}\mathcal{R}} \text{ and } a_y(\mathcal{R}) = \sqrt{\frac{A}{N^2}\frac{1}{\mathcal{R}}}. \tag{A.19}$$

A.4 Energy of the infinite system of point charge molecules

If V_{ij} is the interacting energy of a pair of molecules, the energy of the infinite lattice is:

$$V_{\text{inf}} = \frac{1}{2}\sum_{i\neq j} V_{ij} = \frac{1}{2}\sum_{i=1}^{N_{\text{mol}}}\sum_{j(\neq i)} V_{ij} = \frac{1}{2}\sum_{i=1}^{N_{\text{uc}}}\sum_{i'=1}^{N_{\text{mol/uc}}}\sum_{j(\neq k)} V_{kj}, \tag{A.20}$$

where $k = (i-1)N_{\text{uc}} + i'$, $N_{\text{mol}} = \infty$, $N_{\text{mol/uc}}$ is the number of molecules in the unit cell and N_{uc} is the number of unit cells, assuming there is some periodicity in the infinite system. Now, assuming so, i.e. assuming V_{ij} depends only on the vector $\mathbf{r}_{ij}$ and not $\mathbf{r}_i$ and $\mathbf{r}_j$ separately and assuming the unit cell contains N_{mol} molecules, we can write the energy per unit cell:

$$V_{\text{inf}}/N_{\text{uc}} = \frac{1}{2}\sum_{i=1}^{N_{\text{mol/uc}}}\sum_{j(\neq i)} V_{ij}. \tag{A.21}$$

Here j runs over the infinite set of molecules.

Let's now write the specific expression for the energy. Because it is a 2D problem, it is convenient to use complex numbers instead of vectors [63]. The 2D Coulomb potential energy of two charges sitting at z_1 and z_2 is $-\text{Re}\log(z_1 - z_2)$ [62].

Let's now consider a pair of molecules, with repelling charges at

$$z^a_{1,2} = \zeta_{1,2} + \frac{\Delta_{1,2}}{2}, \tag{A.22}$$

$$z^b_{1,2} = \zeta_{1,2} - \frac{\Delta_{1,2}}{2}. \tag{A.23}$$

With the molecule lengths $|\Delta_1| = |\Delta_2| << |\zeta_{12}|$, where $\zeta_{12} = \zeta_1 - \zeta_2$. Ignoring the interaction within a molecule, the interaction energy is proportional to the real part of

$$\begin{aligned}
&- \log\left[(z^a_1 - z^b_2)(z^b_1 - z^a_2)(z^a_1 - z^a_2)(z^b_1 - z^b_2)\right] = \\
&- 4\log\zeta_{12} - \log\left[1 - \left(\frac{\Delta_1 + \Delta_2}{2\zeta_{12}}\right)^2\right] - \\
&\log\left[1 - \left(\frac{\Delta_1 - \Delta_2}{2\zeta_{12}}\right)^2\right] \\
&= -4\log\zeta_{12} + \frac{\Delta_1^2 + \Delta_2^2}{2\zeta_{12}^2} + \frac{\Delta_1^4 + \Delta_2^4 + 6\Delta_1^2\Delta_2^2}{(2\zeta_{12})^4} \\
&+ \mathcal{O}((\Delta_1/\zeta_{12})^6).
\end{aligned} \tag{A.24}$$

Let's now calculate the energy per unit cell. The first term in the expansion is addressed by putting the molecules in a triangular lattice. Let's start by calculating the next to leading order contribution. We will derive an expression for the case of arbitrary number of molecules in the unit cell and for an arbitrary periodic lattice geometry. In order to follow the calculations let's illustrate one concrete case:

We need to sum the interaction energy of each of the molecules in only one unit cell, with all the other molecules in the infinite lattice. Let's start with the contribution coming from pairs, where one of the molecules is outside of the unit cell. Let ω_1 and ω_2 be the vectors connecting adjacent unit cells and $\{z_{ij}\}$ the vectors connecting molecules within a unit cell, in complex notation [63]. We first do equal color (i.e. equal position in the unit cell) pairs, each of them contributes:

$$\sum_{\substack{n,m \\ n^2+m^2>0}} \frac{(\Delta_i^2 + \Delta_i^2)}{2(n\omega_1 + m\omega_2)^2}. \tag{A.25}$$

Here and from now on, n, m run over all the integers that fulfill the condition $n^2 + m^2 > 0$ (if specified). The indices i, j and k run over molecules inside the unit cell. Therefore, the

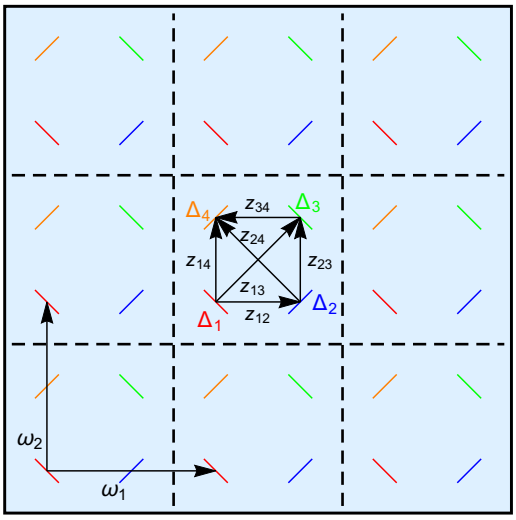


Figure A.1: A schematic square lattice of molecules with $N_{\text{mol}} = 4$. Even though our numerical calculations are done for the triangular lattice, we show a square lattice to emphasise that our result of the general expression of $V_{\text{inf}}/N_{\text{uc}}$ is valid for any lattice geometry.

total contribution is:

$$\sum_{i=1}^{N_{\text{mol}}} \sum_{\substack{n,m \\ n^2+m^2>0}} \frac{(\Delta_i^2 + \Delta_i^2)}{2(n\omega_1 + m\omega_2)^2}. \tag{A.26}$$

The contribution from different color pairs is only slightly trickier, $-z_{ij}$ appears for some of the terms in the denominator, but because of the square, the signs disappear. The contribution of each of the members of the unit cell is:

$$\sum_{j=1}^{N_{\text{mol}}} \sum_{\substack{n,m \\ n^2+m^2>0}} \frac{(\Delta_i^2 + \Delta_j^2)}{2(z_{ij} + n\omega_1 + m\omega_2)^2}. \tag{A.27}$$

Therefore the total contribution is:

$$\sum_{i\neq j}^{N_{\text{mol}}} \sum_{\substack{n,m \\ n^2+m^2>0}} \frac{(\Delta_i^2 + \Delta_j^2)}{2(z_{ij} + n\omega_1 + m\omega_2)^2}. \tag{A.28}$$

The contribution from pairs inside the unit cell is just:

$$\sum_{i\neq j}^{N_{\text{mol}}} \frac{(\Delta_i^2 + \Delta_j^2)}{2z_{ij}^2}. \tag{A.29}$$

Summing all contributions Eq. (A.26), Eq. (A.28) and Eq. (A.29) and including the factor $1/2$ from Eq. (A.21), we get the final expression:

$$V_{\text{inf}}/N_{\text{uc}} = \frac{1}{2}\text{Re}\Bigg[\sum_{i}^{N_{\text{mol}}}\Delta_i^2 \sum_{\substack{n,m \\ n^2+m^2>0}} \frac{1}{(n\omega_1+m\omega_2)^2} + \sum_{i<j}(\Delta_i^2+\Delta_j^2)\sum_{n,m}\frac{1}{(z_{ij}+n\omega_1+m\omega_2)^2}\Bigg]. \tag{A.30}$$

The first term is divergent unless we require the net quadrupole moment of the unit cell to vanish $\sum_i \Delta_i^2 = 0$. Using this condition and due to the double periodicity in z_{ij} of the infinite sum in the second term, $V_{\text{inf}}/N_{\text{uc}} = 0$.

We next calculate the contribution of the third term in the multipole expansion Eq. (A.24). Following the same steps as above we arrive to

$$V_{\text{inf}}/N_{\text{uc}} = \frac{1}{2^4}\text{Re}\Bigg[4\sum_{i}\Delta_i^4 \sum_{\substack{n,m \\ n^2+m^2>0}} \frac{1}{(n\omega_1+m\omega_2)^4} + \sum_{i<j}(\Delta_i^4+\Delta_j^4+6\Delta_i^2\Delta_j^2)\sum_{n,m}\frac{1}{(z_{ij}+n\omega_1+m\omega_2)^4}\Bigg]. \tag{A.31}$$

Using $\sum_i \Delta_i^2 = 0$ ($\Rightarrow \sum_i \Delta_i^4 = -2\sum_{i<j}\Delta_i^2\Delta_j^2$) and taking advantage of the double periodicity in z_{ij} of the last infinite sum, we can rewrite the expression as

$$V_{\text{inf}}/N_{\text{uc}} = \frac{1}{8}\text{Re}\left[\sum_{i<j}\Delta_i^2\Delta_j^2\mathcal{Q}(z_{ij};\omega_1,\omega_2)\right], \tag{A.32}$$

where

$$\begin{aligned}\mathcal{Q}(z_{ij};\omega_1,\omega_2) =& 3\sum_{n,m}\frac{1}{(z_{ij}+n\omega_1+m\omega_2)^4} \\ &- 4\sum_{\substack{n,m \\ n^2+m^2>0}}\frac{1}{(n\omega_1+m\omega_2)^4} \\ &- \sum_{k(\neq a)}\sum_{n,m}\frac{1}{(z_{ak}+n\omega_1+m\omega_2)^4}.\end{aligned} \tag{A.33}$$

Here a is any molecule in the unit cell. $\mathcal{Q}(z_{ij};\omega_1,\omega_2)$ is an elliptic function with periods ω_1 and ω_2. We can numerically calculate it truncating the sums.

A.5 Converting between hyperfine and Cartesian bases

In the hyperfine basis, the Gross-Pitaevskii energy of the spinor BEC is

$$E(\psi) = \int d\mathbf{r} \left(\sum_{\sigma} \left(\frac{|\nabla \psi_\sigma(\mathbf{r})|^2}{2} - \mu \right) + \frac{W_0}{2} \Big(\sum_{\sigma} |\psi_\sigma(\mathbf{r})|^2 \Big)^2 + \frac{W_2}{2} \Big(\sum_{\sigma_1 \sigma_2} \psi^*_{\sigma_1}(\mathbf{r}) \mathbf{F}_{\sigma_1,\sigma_2} \psi_{\sigma_2}(\mathbf{r}) \Big)^2 \right). \tag{A.34}$$

By defining the spinor $\Psi = (\psi_1, \psi_0, \psi_{-1})^T$, the energy functional may be expressed as

$$E(\Psi) = \int d\mathbf{r} \left(\frac{\hbar^2}{2m} \partial_i \Psi^\dagger \cdot \partial_i \Psi - \mu \Psi^\dagger \cdot \Psi + \frac{W_0}{2} (\Psi^\dagger \cdot \Psi)^2 + \frac{W_2}{2} (\Psi^\dagger \mathbf{F} \Psi)^2 \right). \tag{A.35}$$

The spin density is

$$\Psi^\dagger \mathbf{F} \Psi = (\Psi^\dagger F_x \Psi, \Psi^\dagger F_y \Psi, \Psi^\dagger F_z \Psi), \tag{A.36}$$

where

$$F_x = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad F_y = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad F_z = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \tag{A.37}$$

Let us define the spinor in the Cartesian basis as

$$\Upsilon \equiv \begin{pmatrix} \psi_x \\ \psi_y \\ \psi_z \end{pmatrix} = \underbrace{\frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 0 & 1 \\ -i & 0 & -i \\ 0 & \sqrt{2} & 0 \end{pmatrix}}_{M} \Psi. \tag{A.38}$$

The spin density then becomes

$$\Psi^\dagger F_i \Psi = \Psi^\dagger M^{-1} M F_i M^{-1} M \Psi = \Upsilon^\dagger M F_i M^{-1} \Upsilon = \Upsilon^\dagger (-i \boldsymbol{\epsilon}_i) \Upsilon, \tag{A.39}$$

where ϵ_{ijk} is the Levi-Civita symbol, such that

$$\epsilon_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}, \quad \epsilon_y = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \epsilon_z = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{A.40}$$

Note that

$$\Psi^\dagger \cdot \Psi = \Upsilon^\dagger \cdot \Upsilon \quad \text{and} \quad \partial_i \Psi^\dagger \cdot \partial_i \Psi = \partial_i \Upsilon^\dagger \cdot \partial_i \Upsilon. \tag{A.41}$$

Thus,

$$E(\Upsilon) = \int d\mathbf{r} \left(\frac{1}{2} \partial_i \Upsilon^\dagger \cdot \partial_i \Upsilon - \mu \Upsilon^\dagger \cdot \Upsilon + \frac{W_0}{2} (\Upsilon^\dagger \cdot \Upsilon)^2 + \frac{W_2}{2} (\Upsilon^\dagger (-i\boldsymbol{\epsilon}) \Upsilon)^2 \right). \tag{A.42}$$

This can be further simplified by defining $\Upsilon \equiv \mathbf{a} + i\mathbf{b}$, where $\mathbf{a}$ and $\mathbf{b}$ are real vectors. Using the relation $\mathbf{a}\boldsymbol{\epsilon}_i\mathbf{b} = (\mathbf{a} \times \mathbf{b})_i$,

$$\Upsilon^\dagger (-i\boldsymbol{\epsilon_i}) \Upsilon = 2(\mathbf{a} \times \mathbf{b})_i. \tag{A.43}$$

Hence,

$$\begin{aligned} (\Upsilon^\dagger (-i\boldsymbol{\epsilon}) \Upsilon)^2 &= 4(\mathbf{a} \times \mathbf{b})^2 = 4a^2 b^2 - 4(\mathbf{a} \cdot \mathbf{b})^2, \\ (\Upsilon^\dagger \cdot \Upsilon)^2 &= a^4 + b^4 + 2a^2 b^2. \end{aligned} \tag{A.44}$$

The resulting expression for the energy in the Cartesian basis is

$$E(\Upsilon) = \int d\mathbf{r} \left(\frac{1}{2} \partial_i \Upsilon^\dagger \cdot \partial_i \Upsilon - \mu \Upsilon^\dagger \cdot \Upsilon + \frac{W_0 + W_2}{2} (\Upsilon^\dagger \cdot \Upsilon)^2 - \frac{W_2}{2} |\Upsilon \cdot \Upsilon|^2 \right). \tag{A.45}$$

A.6 Interactions between vortices

Parametrizing the wavefunction in terms its amplitude and phase, the Gross-Pitaevskii energy of a scalar BEC is

$$E = \int d\mathbf{r} \left[\frac{1}{2} \left(\nabla \sqrt{\rho(\mathbf{r})} \right)^2 + \frac{\rho(\mathbf{r})}{2} (\nabla \varphi(\mathbf{r}))^2 - \mu \rho(\mathbf{r}) + \frac{U}{2} \rho^2(\mathbf{r}) \right]. \tag{A.46}$$

Let us separate our space into two subregions: the bulk region D where $\rho = \rho_0$ and $\nabla \rho = 0$ and the vortex core regions $\{D_k\}$ where the density decreases from its bulk value ρ_0 to 0. We define the boundary of the domain D as $\partial D = \partial G + \sum_k \partial M_k$, where ∂G is the boundary of the whole system and $\{\partial M_k\}$ are the boundaries of each core region $\{D_k\}$. We can then separate the energy that corresponds to each of the domains as

$$E = E_{\text{core}} + E_D, \quad \text{where} \tag{A.47}$$

$$E_D = A_D\left(-\mu\rho_0 + \frac{U}{2}\rho_0^2\right) + E_\varphi \quad \text{and} \tag{A.48}$$

$$E_\varphi = \frac{\rho_0}{2}\int_D d\mathbf{r}\,(\nabla\varphi(\mathbf{r}))^2. \tag{A.49}$$

From here on we focus on the bulk energy contribution E_φ because as we will see, it is responsible for logarithmic interactions. Using the Helmholtz decomposition of the gradient field $\nabla\varphi$, we can write it as the sum of an irrotational $\nabla\phi$ and divergence free field $\nabla\times(\hat{z}V)$

$$\begin{aligned} E_\varphi = \frac{\rho_0}{2}\int_D d\mathbf{r}\,(\nabla\varphi)^2 &= \frac{\rho_0}{2}\int_D d\mathbf{r}\,(\nabla\phi + \nabla\times(\hat{z}V))^2 \\ &= \frac{\rho_0}{2}\int_D d\mathbf{r}\left[(\nabla\phi)^2 + 2\nabla\phi\cdot\nabla\times(\hat{z}V) + (\nabla\times\hat{z}V)^2\right]. \end{aligned} \tag{A.50}$$

The second term vanishes, since following an integration by parts,

$$\int d\mathbf{r}\,\nabla\phi\cdot\nabla\times(\hat{z}V) = -\int d\mathbf{r}\,\phi\nabla\cdot(\nabla\times\hat{z}V) = 0 \tag{A.51}$$

The third term can be simplified by noting that

$$(\nabla\times\hat{z}V)^2 = (\nabla\times\hat{z}V)\cdot(\nabla\times\hat{z}V) = (\partial_y V, -\partial_x V)\cdot(\partial_y V, -\partial_x V) = (\nabla V)^2. \tag{A.52}$$

Finally Eq. (A.50) becomes,

$$E_\varphi = \frac{\rho_0}{2}\int_D d\mathbf{r}\left[(\nabla\phi)^2 + (\nabla V)^2\right] = E_\phi + E_V. \tag{A.53}$$

The term E_ϕ is the contribution of curl-free fluctuations of the phase, and E_V is the remaining part, which is only non zero in the presence of vortices. Since we are interested in obtaining an expression for the energy of a vortex lattice, we will now focus on the rotational contribution E_V. First, let us derive an expression for V in the presence of a vortex lattice.

In the presence of a vortex lattice, the circulation of $\nabla\varphi$ around the boundary of D (i.e.

along ∂D) is the sum of the contributions from each vortex

$$\oint_{\partial D} \nabla\varphi \cdot d\mathbf{s} = 2\pi \sum_i n_i, \tag{A.54}$$

where n_i are the winding numbers of each of the vortices at positions $\{\mathbf{r}_i\}$. Now, using Stoke's theorem,

$$2\pi \sum_i n_i = \oint_{\partial D} \nabla\varphi \cdot d\mathbf{s} = \int (d\mathbf{r}\hat{z}) \cdot \nabla \times \nabla\varphi, \tag{A.55}$$

and so we can identify

$$\nabla \times \nabla\varphi = 2\pi\hat{z} \sum_i n_i \delta(\mathbf{r} - \mathbf{r}_i). \tag{A.56}$$

Performing the Helmholtz decomposition

$$\nabla \times \nabla\varphi = \nabla \times \nabla\phi + \nabla \times \nabla \times (\hat{z}V) = \hat{z}\nabla^2 V - \nabla \underbrace{(\nabla \cdot \hat{z}V)}_{=0(\nabla \perp \hat{z})} = \hat{z}\nabla^2 V \tag{A.57}$$

and so

$$\nabla^2 V = 2\pi \sum_i n_i \, \delta(\mathbf{r} - \mathbf{r}_i) \quad \Rightarrow \quad V(\mathbf{r}) = \sum_i n_i \, \log(|\mathbf{r} - \mathbf{r}_i|). \tag{A.58}$$

By analogy with electrostatics $\nabla^2 V = -\rho/\epsilon_0$, V can be identified as the potential created by a set of point charges.

Let us now continue with the calculation of E_V in the presence of a vortex lattice.

$$\begin{aligned}
E_V &= \frac{\rho_0}{2} \int_D d\mathbf{r} \, (\nabla V)^2 \\
&= \frac{\rho_0}{2} \int_D d\mathbf{r} \left[\nabla \cdot V\nabla V - V\nabla^2 V\right] \\
&= \frac{\rho_0}{2} \int_D d\mathbf{r} \, \nabla \cdot V\nabla V - \underbrace{\frac{\rho_0}{2} \int_D d\mathbf{r} \left[\Big(\sum_i n_i \, \log(|\mathbf{r} - \mathbf{r}_i|)\Big)\Big(2\pi \sum_j n_j \, \delta(\mathbf{r} - \mathbf{r}_j)\Big)\right]}_{=0,\ \mathbf{r}_j \text{ is out of the integration domain}} \\
&= \frac{\rho_0}{2} \int_D d\mathbf{r} \, \nabla \cdot V\nabla V
\end{aligned} \tag{A.59}$$

Using the divergence theorem

$$\int_D d\mathbf{r} \, \nabla \cdot V\nabla V = \oint_{\partial D = \partial G + \sum_k \partial M_k} V\nabla V \cdot \hat{\mathbf{n}} \, ds, \tag{A.60}$$

where ∂D is the boundary of the domain D with positive orientation (always leaving D on the left while walking along ∂D). Since we have a non-simply connected domain due to the separation of the vortex regions, $\partial D = \partial G + \sum_k \partial M_k$, where ∂G is the outer boundary of the entire system while $\sum_k \partial M_k$ is the sum of the paths around each vortex. Positive orientation corresponds to ∂G being counter clockwise whereas each ∂M_k being clockwise. For convenience, we will proceed with the calculation in the complex plane, parameterized by $z = x + iy$. Then,

$$\partial_z = \frac{1}{2}(\partial_x - i\partial_y), \tag{A.61}$$

$$d\mathbf{n} = (dy, -dx), \tag{A.62}$$

$$\nabla V \cdot d\mathbf{n} = (\partial_x V dy - \partial_y V dx) = \frac{1}{i}(\partial_z V dz - \partial_{\bar{z}} V d\bar{z}), \tag{A.63}$$

$$V = \sum_i n_i \log|z - z_i| = \sum_i \frac{n_i}{2}(\log(z - z_i) + \log(\bar{z} - \bar{z}_i)), \tag{A.64}$$

$$\partial_z V = \sum_i \frac{n_i}{2}\frac{1}{(z - z_i)} \quad , \quad \partial_{\bar{z}} V = \sum_i \frac{n_i}{2}\frac{1}{(\bar{z} - \bar{z}_i)}, \tag{A.65}$$

$$\begin{aligned}
\oint_{\sum_k \partial M_k} V\nabla V \cdot \hat{\mathbf{n}}\, ds &= \frac{1}{2i}\sum_{ij} n_i n_j \Big(\oint_{\sum_k \partial M_k} \frac{\log|z - z_j|}{(z - z_i)}\, dz - \oint_{\sum_k \partial M_k} \frac{log|z - z_j|}{(\bar{z} - \bar{z}_i)}\, d\bar{z} \Big) \\
&= \frac{1}{2i}\sum_{ijk} n_i n_j \Big(\oint_{\partial M_k} \frac{\log|z - z_j|}{(z - z_i)}\, dz - \oint_{\partial M_k} \frac{\log|z - z_j|}{(\bar{z} - \bar{z}_i)}\, d\bar{z} \Big) \\
&\simeq \frac{1}{2i}\sum_{ijk} n_i n_j \log|z_k - z_j| \Big(\oint_{\partial M_k} \frac{dz}{(z - z_i)} - \oint_{\partial M_k} \frac{d\bar{z}}{(\bar{z} - \bar{z}_i)} \Big) \\
&= \frac{1}{2i}\sum_{ijk} n_i n_j \log|z_k - z_j| \Big(\oint_{\partial M_k} \frac{dz}{(z - z_i)} - \oint_{\overline{\partial M_k}} \frac{dz}{(z - \bar{z}_i)} \Big) \\
&= \frac{1}{2i}\sum_{ij} n_i n_j \log|z_i - z_j| \Big(-2\pi i - 2\pi i \Big) \\
&= -2\pi \sum_{ij} n_i n_j \log|z_i - z_j|,
\end{aligned} \tag{A.66}$$

where we used the Cauchy formula to evaluate the integrals. Note that in the Cauchy formula, the path should be counter clockwise and therefore we pick up a sign from integrating over ∂M_k. On the other hand, $\overline{\partial M_k}$ is counter clockwise so there is no extra

sign from using the Cauchy formula. In the third line of Eq. (A.66), the approximation is valid in the point-like vortex regime. From this expression we see that there is a $2D$ Coulomb interaction between the vortices, n_i being the charge of the vortex at position z_i

$$E_{\text{int}} = -\pi\rho_0 \sum_{ij} n_i n_j \log|z_i - z_j| = E'_{\text{core}} - 2\pi\rho_0 \sum_{i<j} n_i n_j \log|z_i - z_j|, \tag{A.67}$$

where we throw the self-interaction to the core contribution that we disregard for the derivation of logarithmic interactions. For the integral over ∂G, if we take our system to be infinitely large, $V = \sum_i n_i \log|z - z_i| \sim \log|z| \sum_i n_i$. Thus,

$$\begin{aligned} E_{\text{charge}} = \frac{\rho_0}{2} \oint_{\partial G} V\nabla V \cdot \hat{\mathbf{n}}\, ds &\simeq \frac{\rho_0}{2i} \oint_{\partial G} \frac{\log|z|}{z}\, dz \sum_{ij} n_i n_j \\ &\simeq \pi\rho_0 \left(\sum_i n_i\right)^2 \log\left(\frac{L}{\xi}\right). \end{aligned} \tag{A.68}$$

In the case where there is no net charge $\sum_i n_i = 0$, this term vanishes.

On the other hand it is not a very realistic approximation. At least in the case where all vortices have the same sign in the scalar BEC, the repulsion will separate the vortices by a distance comparable to the system size. Then $V = \sum_i n_i \log|z - z_i| \nsim \log|z| \sum_i n_i$ and so

$$\begin{aligned} E_{\text{charge}} &= \frac{\rho_0}{2} \oint_{\partial G} V\nabla V \cdot \hat{\mathbf{n}}\, ds \\ &= \frac{\rho_0}{4i} \sum_{ij} n_i n_j \left[\oint_{\partial G} \frac{\log|z - z_i|}{(z - z_j)} dz - \oint_{\partial G} \frac{\log|\bar{z} - \bar{z}_i|}{(\bar{z} - \bar{z}_j)} d\bar{z}\right]. \end{aligned} \tag{A.69}$$

Putting together all the energy contributions, the energy of the vortex lattice takes the form

$$E = A_D \left(-\mu\rho_0 + \frac{U}{2}\rho_0^2\right) + E_{\text{core}} + E'_{\text{core}} + E_\phi + E_{\text{int}} + E_{\text{charge}}. \tag{A.70}$$

A.6.1 Charge energy of a pair of vortices on a disk geometry

Let us calculate the charge energy of a pair of vortices in a disk. The winding numbers are n_1, n_2 and the pair sits in the real axis at positions $\pm r$. The charge energy is

$$E_{\text{charge}} = \frac{\rho_0}{2i}\oint_{|z|=R}\left[n_1^2\frac{\log|z-r|}{(z-r)}dz + n_2^2\frac{\log|z+r|}{(z+r)}dz + n_1n_2\frac{\log|z-r|}{(z+r)}dz + n_1n_2\frac{\log|z+r|}{(z-r)}dz\right], \tag{A.71}$$

where we used the relation

$$\oint_{|z|=R}\frac{\log|\bar{z}-r|}{(\bar{z}-r)}d\bar{z} = -\oint_{|z|=R}\frac{\log|z-r|}{(z-r)}dz. \tag{A.72}$$

Furthermore, for any $\omega \in \mathbb{C}$ where $|\omega| < R$

$$\begin{aligned}\oint_{|z|=R}\frac{\log|z-\omega|}{(z-\omega)}dz &= i\pi\log\left(R^2-|\omega|^2\right), \\ \oint_{|z|=R}\frac{\log|z+\omega|}{(z-\omega)}dz &= i\pi\log\left(R^2+|\omega|^2\right).\end{aligned} \tag{A.73}$$

Finally,

$$E_{\text{charge}} = \frac{\pi\rho_0}{2}\left[\left(n_1^2+n_2^2\right)\log\left(R^2-r^2\right) + 2n_1n_2\log\left(R^2+r^2\right)\right]. \tag{A.74}$$

B

Gaussian and Poisson processes

B.1 Relation between covariance functions and SDEs

Most of the theory explained in this appendix can be found in Chapters 6 and 12 of Ref. [123]. A stochastic differential equation (SDE) is a differential equation that contains terms which are random functions. This implies that their solutions are also random functions. Consider a Gaussian noise-driven ordinary differential equation of the form

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + \mathbf{L}(\mathbf{x}, t)d\boldsymbol{\beta}(t), \tag{B.1}$$

where $\boldsymbol{\beta}(t)$ is Brownian motion with diffusion matrix $\mathbf{Q}$ and $\mathbf{f}(\mathbf{x}, t)$ and $\mathbf{L}(\mathbf{x}, t)$ are arbitrary vector- and matrix-valued functions, respectively. The solutions $\mathbf{x}(t)$ of SDEs are random processes and therefore they have certain probability distribution $p(\mathbf{x}(t))$ [also denoted $p(\mathbf{x}, t)$]. This probability density solves the *Fokker–Planck–Kolmogorov* (FPK) equation

$$\begin{aligned}\frac{\partial p(\mathbf{x}, t)}{\partial t} = &- \sum_i \frac{\partial}{\partial x_i}\left[f_i(\mathbf{x}, t)p(\mathbf{x}, t)\right] \\ &+ \frac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial x_i \partial x_j}\left\{\left[\mathbf{L}(\mathbf{x}, t)\mathbf{Q}\mathbf{L}^T(\mathbf{x}, t)\right]_{ij} p(\mathbf{x}, t)\right\},\end{aligned} \tag{B.2}$$

given the initial condition $p(\mathbf{x}, t_0)$. One can obtain the equations of motion for the mean, covariance and other statistical quantities from this equation. Among others, denoting the mean $\mathbf{m}(t) = \mathbb{E}[\mathbf{x}(t)]$ and the *marginal covariance* $\mathbf{C}(t, t) \equiv \mathbf{P}(t) = \mathbb{E}\left[(\mathbf{x}(t) - \mathbf{m}(t))\,(\mathbf{x}(t) - \mathbf{m}(t))^T\right]$,

$$\frac{d\mathbf{m}}{dt} = \mathbb{E}\left[\mathbf{f}(\mathbf{x}, t)\right], \tag{B.3}$$

$$\frac{d\mathbf{P}}{dt} = \mathbb{E}\left[\mathbf{f}(\mathbf{x}, t)(\mathbf{x} - \mathbf{m})^T\right] + \mathbb{E}\left[(\mathbf{x} - \mathbf{m})\mathbf{f}^T(\mathbf{x}, t)\right] + \mathbb{E}\left[\mathbf{L}(\mathbf{x}, t)\mathbf{Q}\mathbf{L}^T(\mathbf{x}, t)\right]. \tag{B.4}$$

Another useful quantity that we can obtain from the FPK equation (B.2) is the *transition density* $p(\mathbf{x}(t)|\mathbf{x}(s))$ of the SDE in (B.1), which is the probability of the random process taking the value $\mathbf{x}(t)$ at time t, given the value at time s was $\mathbf{x}(s)$. This quantity is the solution of the FPK equation (B.2), with the initial condition $p(\mathbf{x}(t)|\mathbf{x}(s)) = \delta(\mathbf{x}(t) - \mathbf{x}(s))$ at $t = s$.

An SDE is *linear* if $\mathbf{f} = \mathbf{F}\mathbf{x}$. The covariance function $\mathbf{C}(t, t')$ of linear stochastic differential equations can be obtained from the marginal covariance

$$\mathbf{C}(t, t') = \begin{cases} \mathbf{P}(t) \exp\left[(t' - t)\mathbf{F}\right]^T, & \text{if } t < t', \\ \exp\left[(t - t')\mathbf{F}\right] \mathbf{P}(t'), & \text{if } t \geq t'. \end{cases} \tag{B.5}$$

B.1.1 Equivalent discretisations of linear time-invariant SDEs

An SDE is *time-invariant* if $\mathbf{f}$ and $\mathbf{L}$ do not depend on time. Consider the linear time-invariant stochastic differential equation

$$d\mathbf{x} = \mathbf{F}\mathbf{x}dt + \mathbf{L}d\boldsymbol{\beta}, \tag{B.6}$$

with initial conditions $\mathbf{x}(t_0) \sim \mathrm{N}(\mathbf{m}_0, \mathbf{P}_0)$, where $\mathrm{N}(\mathbf{m}_0, \mathbf{P}_0)$ denotes a Gaussian distribution with mean $\mathbf{m}_0$ and marginal covariance $\mathbf{P}_0$. From the FPK equation, one obtains the transition density

$$p\left(\mathbf{x}(t)|\mathbf{x}(s)\right) = N\left(\mathbf{m}(t|s), \mathbf{P}(t|s)\right), \tag{B.7}$$

where

$$\mathbf{m}(t|s) = \exp\left(\mathbf{F}(t - s)\right) \mathbf{x}(s), \tag{B.8}$$

$$\mathbf{P}(t|s) = \int_s^t \exp\left(\mathbf{F}(t - \tau)\right) \mathbf{L}\mathbf{Q}\mathbf{L}^T \exp\left(\mathbf{F}(t - \tau)\right)^T d\tau. \tag{B.9}$$

Let us consider discrete times $\{t_k\}$, separated by Δt_k. Eq. (B.7) then implies

$$\mathbf{x}(t_{k+1}) - \mathbf{m}(t_{k+1}|t_k) = \mathbf{q}_k, \quad \mathbf{q}_k \sim N(0, \mathbf{P}(t_{k+1}|t_k)). \tag{B.10}$$

Therefore, we derive a discrete stochastic equation

$$\mathbf{x}(t_{k+1}) = \exp\left(\mathbf{F}(\Delta t_k)\right) \mathbf{x}(t_k) + \mathbf{q}_k, \quad \mathbf{q}_k \sim N(0, \mathbf{P}(\Delta t_k|0)). \tag{B.11}$$

This discretization is exact in that the probability distribution of the continuous and discrete models defined in Eqs. (B.1) and (B.11), coincide at times $\{t_k\}$.

B.1.2 From steady state covariance functions to discrete stochastic processes

As shown at the beginning of Appendix B.1, it is possible to derive the covariance function of an SDE. Conversely, it is also possible to find the SDE that corresponds to a given covariance function. Consider the following steady state covariance

$$\begin{aligned} &\mathbf{C}(\tau) = \mathbf{C}_{\cos}(\tau)\mathbf{C}_{\exp}(\tau), \text{ where} \\ &\mathbf{C}_{\cos}(\tau) = \cos(\omega\tau), \\ &\mathbf{C}_{\exp}(\tau) = \sigma^2 e^{-\lambda|\tau|}. \end{aligned} \tag{B.12}$$

As shown in Refs. [126, 127], the corresponding SDE is

$$\begin{aligned} d\mathbf{g}(t) &= \mathbf{F}\mathbf{g}(t)dt + \mathbf{L}d\boldsymbol{\beta}, \\ x(t) &= \mathbf{H}\mathbf{g}(t), \end{aligned} \tag{B.13}$$

where

$$\mathbf{F} = \begin{pmatrix} -\lambda & -\omega \\ \omega & -\lambda \end{pmatrix}, \tag{B.14}$$

$$\mathbf{L} = \mathbb{1}_2, \tag{B.15}$$

$$\mathbf{H} = (1, 0). \tag{B.16}$$

This is called a *continuous state space model*, the vector $\mathbf{H}$ is the *measurement model* and $\mathbf{g}(t)$ the *state*. The equivalent discretization of Eq. (B.13) is

$$\mathbf{g}_{k+1} = \mathbf{A}\mathbf{g}_k + \mathbf{q}_k, \quad \mathbf{q}_k \sim N(0, \boldsymbol{\Sigma}), \tag{B.17}$$

$$x(t_k) = \mathbf{H}\mathbf{g}_k, \tag{B.18}$$

where

$$\mathbf{A} = \exp(-\lambda\Delta t)\begin{pmatrix}\cos(\omega\Delta t) & -\sin(\omega\Delta t)\\ \sin(\omega\Delta t) & \cos(\omega\Delta t)\end{pmatrix}, \tag{B.19}$$

$$\mathbf{\Sigma} = \sigma^2(1 - e^{-2\lambda\Delta t})\mathbb{1}_2. \tag{B.20}$$

When the kernel is a sum of N stationary kernels $\mathbf{C}(\tau) = \sum_{i=1}^{N} \mathbf{C}_i(\tau)$, we get the corresponding SDE by replacing $\mathbf{F}, \mathbf{L}$ and $\mathbf{H}$ in Eq. (B.13) by

$$\mathbf{F} = \text{blkdiag}(\mathbf{F}_1, ..., \mathbf{F}_N), \tag{B.21}$$

$$\mathbf{L} = \text{blkdiag}(\mathbf{L}_1, ..., \mathbf{L}_N), \tag{B.22}$$

$$\mathbf{Q} = \text{blkdiag}(\mathbf{Q}_1, ..., \mathbf{Q}_N), \tag{B.23}$$

$$\mathbf{H} = (\mathbf{H}_1, ..., \mathbf{H}_N). \tag{B.24}$$

Here $\mathbf{Q}$ is the diffusion matrix of Brownian motion $\boldsymbol{\beta}(t)$. The corresponding equivalent discretization is obtained by performing the equivalent substitution of $\mathbf{A}, \mathbf{\Sigma}$ and $\mathbf{H}$ in Eq. (B.17). The dimension of the state vector $\mathbf{g}$ in Eqs. (B.13) and (B.17) is then increased by a factor of N.

Obtaining samples from these discrete stochastic processes is equivalent to sampling the corresponding multidimensional Gaussian distributions.

B.2 Filtered Poisson processes

B.2.1 Poisson process

A *standard Poisson process* N_t is a counting process that has jumps of size +1 at homogeneously distributed random times and its path is constant in between two jumps. This is defined as

$$N_t = \sum_{k=1}^{\infty} \mathbb{1}_{[t_k,\infty)}, \quad \text{for } t \geq 0, \tag{B.25}$$

where

$$\mathbb{1}_{[t_k,\infty)} = \begin{cases}1, & \text{if } t \geq t_k,\\ 0, & \text{if } 0 \leq t < t_k.\end{cases} \tag{B.26}$$

Furthermore, a Poisson process satisfies the following conditions:

1. Independence of increments: for all $0 \leq t_0 < t_1 < \cdots < t_n$, the increments

$$N_{t_1} - N_{t_0}, ..., N_{t_n} - N_{t_n - 1}, \tag{B.27}$$

   are independent random variables.

2. Stationarity of increments: $N_{t+h} - N_{s+h}$ and $N_t - N_s$ have the same distribution for all $h > 0$ and $0 \leq s \leq t$.

3. Conditions 1 and 2 imply that the probability distribution of the increments is a Poisson distribution, i.e. for all $0 \leq s \leq t$,

$$p(N_t - N_s = k) = e^{-\lambda(t-s)} \frac{(\lambda(t-s))^k}{k!}. \tag{B.28}$$

   The parameter λ is called the intensity of the Poisson process.

From the last condition we can infer the sort time asymptotics

$$\begin{aligned} p(N_{\Delta t} = 0) &= e^{-\Delta t \lambda} = 1 - \Delta t\lambda + \mathcal{O}(\Delta t^2) \approx 1 - \Delta t\lambda, \\ p(N_{\Delta t} = 1) &= \Delta t \lambda e^{-\Delta t \lambda} = \Delta t\lambda + \mathcal{O}(\Delta t^2) \approx \Delta t\lambda, \quad \Delta t \to 0. \end{aligned} \tag{B.29}$$

B.2.2 One-time characteristic function of a filtered Poisson process

A *filtered Poisson process* (FPP) $X(t)$ consists of the superposition of uncorrelated pulses $\varphi(t - t_k)$, where the arrival times $\{t_k\}$ follow a Poisson distribution

$$X(t) = \sum_k A_k \varphi(t - t_k). \tag{B.30}$$

The overall amplitude A_k is random. Let us consider the case where at each time, A_k can independently take the values $\pm A$, with equal probabilities. The characteristic function of $X(t)$ is

$$\Phi_X(u, t) = \mathbb{E}\left(e^{iuX(t)}\right), \tag{B.31}$$

where the average is taken over all possible Poisson processes. Note that this involves averaging over the random set of jump times $\{t_k\}$ as well as the value of the sequence of amplitudes $\{A_k\}$ at times $\{t_k\}$. Using Campbell's theorem [128],

$$\mathbb{E}\left(e^{iuX(t)}\right) = \exp\left\{-\lambda \int_0^t d\alpha \left[1 - \Phi_A(iu\varphi(t-\alpha))\right]\right\}, \text{ where} \tag{B.32}$$

$$\Phi_A(iu\varphi(t-\alpha)) = \mathbb{E}_A\left(e^{iuA\varphi(t-\alpha)}\right). \tag{B.33}$$

Naive proof:

Let us rewrite the process as a sum over all time steps $\{i\Delta t\}$, instead of the jump times $\{t_k\}$ only

$$X(t) = \sum_{k\in\{t_k\}} A_k \varphi\left(t - t_k\right) = \sum_{i=1}^{N} \sigma_i A_i \varphi\left(t - i\Delta t\right), \tag{B.34}$$

where $t = N\Delta t$. The parameter σ_i is 1 if there is a jump, and 0 otherwise. The expectation value in (B.31) is an average over the random variables A and σ at each time step, i.e.,

$$\Phi_X(u,t) = \mathbb{E}\left(e^{iuX(t)}\right) = \mathbb{E}_\sigma\left[\mathbb{E}_A\left(e^{iuX(t)}\right)\right]. \tag{B.35}$$

Let us define

$$\Phi_{A_i}\left[u\varphi(t - i\Delta t)\right] \equiv \mathbb{E}_{A_i}\left(e^{iu\sigma_i A_i \varphi(t-i\Delta t)}\right). \tag{B.36}$$

Then expectation value at time $t_i = i\Delta t$ is

$$\mathbb{E}\left(e^{iu\sigma_i A_i \varphi(t-i\Delta t)}\right) = \mathbb{E}_{\sigma_i}\left\{\Phi_{A_i}\left[u\varphi(t - i\Delta t)\right]\right\}. \tag{B.37}$$

The probability for there being a jump (i.e., $\sigma = 1$) is $\lambda\Delta t$ and the probability of no jump (i.e., $\sigma = 0$) is $1 - \lambda\Delta t$. Then

$$\mathbb{E}\left(e^{iu\sigma_i A_i \varphi(t-i\Delta t)}\right) = 1 + \lambda\Delta t\left(\Phi_{A_i}\left[u\varphi(t - i\Delta t)\right] - 1\right) \approx \exp\left(-\lambda\Delta t\left\{1 - \Phi_{A_i}\left[u\varphi(t - i\Delta t)\right]\right\}\right). \tag{B.38}$$

The product over all time steps yields (B.32). If the amplitude outcomes are $\pm A$ with equal probabilities, then

$$\mathbb{E}\left(e^{iuX(t)}\right) = \exp\left\{-\lambda \int_0^t d\alpha \left[1 - \cos\left(uA\varphi(t-\alpha)\right)\right]\right\}. \tag{B.39}$$

B.2.3 Two-time characteristic function

The two-time characteristic function is

$$\Phi_x(u_1, t_1; u_2, t_2) = \mathbb{E}\left(e^{i[u_1 X(t_1) + u_2 X(t_2)]}\right). \tag{B.40}$$

Let us consider the case where $t_2 > t_1$. Then $X(t_1)$ depends on all the jumps before t_1 and $X(t_2)$ depends on all jumps before t_2, which includes all those that contributed to $X(t_1)$: this is the source of correlation between the two variables. If we split up $X(t_2)$ as

$$X(t_2) = \sum_{\{k:t_1<t_k<t_2\}} A_k \varphi(t_2 - t_k) + \sum_{\{k:t_1>t_k\}} A_k \varphi(t_2 - t_k), \tag{B.41}$$

the exponent in Eq. (B.40) is expressed as the sum of independent quantities

$$u_1 X(t_1) + u_2 X(t_2) = \sum_{\{k:t_k<t_1\}} A_k \left[u_1 \varphi(t_1 - t_k) + u_2 \varphi(t_2 - t_k)\right] + \sum_{\{k:t_1<t_k<t_2\}} u_2 A_k \varphi(t_2 - t_k). \tag{B.42}$$

Therefore, the characteristic function becomes

$$\Phi_x(u_1, t_1; u_2, t_2) = \mathbb{E}\left(e^{i \sum_{\{k:t_1>t_k\}} A_k \left[u_1 \varphi(t_1 - t_k) + u_2 \varphi(t_2 - t_k)\right]}\right) \mathbb{E}\left(e^{i \sum_{\{k:t_1<t_k<t_2\}} u_2 A_k \varphi(t_2 - t_k)}\right). \tag{B.43}$$

Following the same procedure we followed to derive the one-time characteristic function, we arrive to

$$\begin{aligned}\Phi_X(u_1, t_1; u_2, t_2) = \exp\Bigg\{ &- \lambda \int_{t_1}^{t_2} \left[1 - \Phi_A(i u_2 \varphi(t_2 - \alpha))\right] d\alpha \\ &- \lambda \int_0^{t_1} \left[1 - \Phi_A(i u_1 \varphi(t_1 - \alpha) + i u_2 \varphi(t_2 - \alpha))\right] d\alpha \Bigg\}.\end{aligned} \tag{B.44}$$

If the amplitude outcomes are $\pm A$ with equal probabilities,

$$\begin{aligned}\Phi_X(u_1, t_1; u_2, t_2) = \exp\Bigg\{ &- \lambda \int_{t_1}^{t_2} \left[1 - \cos\left(u_2 \varphi(t_2 - \alpha)\right)\right] d\alpha \\ &- \lambda \int_0^{t_1} \left[1 - \cos\left(u_1 \varphi(t_1 - \alpha) + u_2 \varphi(t_2 - \alpha)\right)\right] d\alpha \Bigg\}.\end{aligned} \tag{B.45}$$

The correlations arise from the second term. For stationary correlations the lower limit should be taken to $-\infty$. It's clear that the coefficients of $u_1 u_2^3$ and $u_2 u_1^3$ will be different, which indicates the absence of time reversal invariance. By Taylor expanding the characteristic function, we have access to all two-time correlators. Specifically,

$$\mathbb{E}\left(X(t_1)X^3(t_2)\right) = \lambda I_{1,3}^{-\infty,t_1} + 3\lambda^2 I_{1,1}^{-\infty,t_1}\left(I_{0,2}^{-\infty,t_1} + I_{0,2}^{t_1,t_2}\right), \tag{B.46}$$

$$\mathbb{E}\left(X^3(t_1)X(t_2)\right) = \lambda I_{3,1}^{-\infty,t_1} + 3\lambda^2 I_{1,1}^{-\infty,t_1} I_{2,0}^{-\infty,t_1}, \text{ where} \tag{B.47}$$

$$I_{n,m}^{t,t'} = \int_t^{t'} d\alpha\, \varphi^n(t_1 - \alpha)\varphi^m(t_2 - \alpha). \tag{B.48}$$

This result is general for any characteristic function of the form shown in Eq. (B.45), taking the lower limit to $-\infty$.

Bibliography

[1] "Onsager's quantization of circulation in superfluid helium," in *The Collected Works of Lars Onsager*, pp. 693–728.

[2] L. Onsager, Nuovo Cimento (Suppl.) **6**, 249 (1949).

[3] F. London, *Superfluids: Vol. I.* (Wiley, 1950).

[4] R. Feynman, Progress in Low Temperature Physics, **1**, 17 (1955).

[5] A. A. Abrikosov, Sov. Phys. JETP **5** (1957).

[6] W. Vinen, Nature **181**, 1524 (1958).

[7] B. S. Deaver Jr and W. M. Fairbank, Physical Review Letters **7**, 43 (1961).

[8] M. H. Anderson, J. R. Ensher, M. R. Matthews, C. E. Wieman, and E. A. Cornell, Science **269**, 198 (1995), https://science.sciencemag.org/content/269/5221/198.full.pdf .

[9] A. L. Fetter, Reviews of Modern Physics **81**, 647 (2009).

[10] S. Stock, B. Battelier, V. Bretin, Z. Hadzibabic, and J. Dalibard, Laser Physics Letters **2**, 275 (2005).

[11] F. London, Phys. Rev. **54**, 947 (1938).

[12] C. J. Pethick and H. Smith, *Bose-Einstein condensation in dilute gases* (Cambridge university press, 2002) p. 161.

[13] C. J. Myatt, E. A. Burt, R. W. Ghrist, E. A. Cornell, and C. E. Wieman, Phys. Rev. Lett. **78**, 586 (1997).

[14] D. M. Stamper-Kurn, M. R. Andrews, A. P. Chikkatur, S. Inouye, H.-J. Miesner, J. Stenger, and W. Ketterle, Phys. Rev. Lett. **80**, 2027 (1998).

[15] R. Ors, Annals of Physics **349**, 117 (2014).

[16] J. I. Cirac and F. Verstraete, Journal of Physics A: Mathematical and Theoretical **42**, 504004 (2009).

[17] T. Xiang, J. Lou, and Z. Su, Phys. Rev. B **64**, 104414 (2001).

[18] U. Schollwck, Annals of Physics **326**, 96 (2011), january 2011 Special Issue.

[19] F. Verstraete and J. I. Cirac, Phys. Rev. Lett. **104**, 190405 (2010).

[20] C. Schön, E. Solano, F. Verstraete, J. I. Cirac, and M. M. Wolf, Phys. Rev. Lett. **95**, 110503 (2005).

[21] N. Buduma and N. Locascio, *Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms*, 1st ed. (O'Reilly Media, Inc., 2017).

[22] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016) http://www.deeplearningbook.org.

[23] F. Rosenblatt, Psychological Review , 65 (1958).

[24] Wikipedia contributors, "Automatic differentiation — Wikipedia, the free encyclopedia," (2019), [Online; accessed 8-July-2019].

[25] D. P. Kingma and P. Dhariwal, in *Advances in Neural Information Processing Systems 31*, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Curran Associates, Inc., 2018) pp. 10215–10224.

[26] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman, Behavioral and Brain Sciences **40**, e253 (2017).

[27] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, in *Advances in Neural Information Processing Systems 27*, edited by Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (Curran Associates, Inc., 2014) pp. 2672–2680.

[28] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," (2013), arXiv:1312.6114 .

[29] B. J. Frey, *Graphical Models for Machine Learning and Digital Communication* (MIT Press, Cambridge, MA, USA, 1998) p. 59.

[30] Y. Bengio and S. Bengio, in *Proceedings of the 12th International Conference on Neural Information Processing Systems*, NIPS'99 (MIT Press, Cambridge, MA, USA, 1999) pp. 400–406.

[31] L. Dinh, D. Krueger, and Y. Bengio, "Nice: Non-linear independent components estimation," (2014), arXiv:1410.8516 .

[32] V. Tkachenko, Sov. Phys. JETP **22**, 1282 (1966).

[33] L. J. Campbell, in *Mathematical Aspects of Vortex Dynamics* (1989) pp. 195–204.

[34] W. H. Kleiner, L. M. Roth, and S. H. Autler, Phys. Rev. **133**, A1226 (1964).

[35] E. Brandt, physica status solidi (b) **51**, 345 (1972).

[36] M. M. Salomaa and G. E. Volovik, Rev. Mod. Phys. **59**, 533 (1987).

[37] D. M. Stamper-Kurn and M. Ueda, Rev. Mod. Phys. **85**, 1191 (2013).

[38] D. T. Son and M. A. Stephanov, Phys. Rev. A **65**, 063621 (2002).

[39] M. Tylutki, L. P. Pitaevskii, A. Recati, and S. Stringari, Physical Review A **93**, 043623 (2016).

[40] C. Qu, M. Tylutki, S. Stringari, and L. P. Pitaevskii, Physical Review A **95**, 033614 (2017).

[41] L. Calderaro, A. L. Fetter, P. Massignan, and P. Wittek, Phys. Rev. A **95**, 023605 (2017).

[42] J. J. García-Ripoll, V. M. Pérez-García, and F. Sols, Phys. Rev. A **66**, 021602 (2002).

[43] K. Kasamatsu, M. Tsubota, and M. Ueda, Physical review letters **93**, 250406 (2004).

[44] M. Cipriani and M. Nitta, Physical review letters **111**, 170401 (2013).

[45] A. Aftalion and P. Mason, Physical Review A **94**, 023616 (2016).

[46] M. Salomaa and G. Volovik, Physical review letters **55**, 1184 (1985).

[47] S. Autti, V. Dmitriev, J. Mäkinen, A. Soldatov, G. Volovik, A. Yudin, V. Zavjalov, and V. Eltsov, Physical review letters **117**, 255301 (2016).

[48] T.-L. Ho, Phys. Rev. Lett. **87**, 060403 (2001).

[49] N. R. Cooper, Advances in Physics **57**, 539 (2008).

[50] L. Mingarelli, E. E. Keaveny, and R. Barnett, Journal of Physics: Condensed Matter **28**, 285201 (2016).

[51] T. Kita, T. Mizushima, and K. Machida, Phys. Rev. A **66**, 061601 (2002).

[52] J. W. Reijnders, F. J. M. van Lankvelt, K. Schoutens, and N. Read, Phys. Rev. A **69**, 023612 (2004).

[53] E. J. Mueller and T.-L. Ho, Phys. Rev. Lett. **88**, 180403 (2002).

[54] M. Keçeli and M. O. Oktel, Phys. Rev. A **73**, 023611 (2006).

[55] M. O. Scully and M. S. Zubairy, *Quantum-optics* (Cambridge university press, 1999) pp. 148–152.

[56] C. Cohen-Tannoudji, B. Diu, and F. Laloë, *Quantum mechanics*, Vol. 1 (WILEY-VCH, 1977) p. 407.

[57] M. R. Matthews, B. P. Anderson, P. C. Haljan, D. S. Hall, M. J. Holland, J. E. Williams, C. E. Wieman, and E. A. Cornell, Phys. Rev. Lett. **83**, 3358 (1999).

[58] T. Kita, Journal of the Physical Society of Japan **67**, 2067 (1998).

[59] L. Mingarelli, E. E. Keaveny, and R. Barnett, "Vortex lattices in binary mixtures of repulsive superfluids," (2017), arXiv:1712.04882 .

[60] E. Jones, T. Oliphant, P. Peterson, *et al.*, "SciPy: Open source scientific tools for Python," (2001–), [Online; accessed ¡today¿].

[61] A. Lamacraft, Physical Review A **77**, 063622 (2008).

[62] M. Kardar, *Statistical Physics of Fields* (Cambridge University Press, 2007) p. 167.

[63] L. Milne-Thompson, *Theoretical hydrodynamics* (Macmillan Press, 1979) p. 121.

[64] T.-L. Ho and V. B. Shenoy, Phys. Rev. Lett. **77**, 3276 (1996).

[65] E. Timmermans, Phys. Rev. Lett. **81**, 5718 (1998).

[66] B. Van Schaeybroeck, Phys. Rev. A **78**, 023624 (2008).

[67] K. KASAMATSU, M. TSUBOTA, and M. UEDA, International Journal of Modern Physics B **19**, 1835 (2005), https://doi.org/10.1142/S0217979205029602 .

[68] G. W. Semenoff and F. Zhou, Phys. Rev. Lett. **98**, 100401 (2007).

[69] M. Kobayashi, Y. Kawaguchi, M. Nitta, and M. Ueda, Phys. Rev. Lett. **103**, 115301 (2009).

[70] H. T. C. Stoof, E. Vliegen, and U. Al Khawaja, Phys. Rev. Lett. **87**, 120407 (2001).

[71] J.-P. Martikainen, A. Collin, and K.-A. Suominen, Phys. Rev. Lett. **88**, 090404 (2002).

[72] C. M. Savage and J. Ruostekoski, Phys. Rev. A **68**, 043604 (2003).

[73] T.-L. Ho, Phys. Rev. Lett. **81**, 742 (1998).

[74] U. Leonhardt and G. E. Volovik, Journal of Experimental and Theoretical Physics Letters **72**, 46 (2000).

[75] A. E. Leanhardt, Y. Shin, D. Kielpinski, D. E. Pritchard, and W. Ketterle, Phys. Rev. Lett. **90**, 140403 (2003).

[76] J. Ruostekoski and J. R. Anglin, Phys. Rev. Lett. **86**, 3934 (2001).

[77] R. A. Battye, N. R. Cooper, and P. M. Sutcliffe, Phys. Rev. Lett. **88**, 080401 (2002).

[78] C. M. Savage and J. Ruostekoski, Phys. Rev. Lett. **91**, 010403 (2003).

[79] U. A. Khawaja and H. T. C. Stoof, Phys. Rev. A **64**, 043612 (2001).

[80] M. A. Metlitski and A. R. Zhitnitsky, Journal of High Energy Physics **2004**, 017 (2004).

[81] J. Ruostekoski, Phys. Rev. A **70**, 041601 (2004).

[82] Y. Kawaguchi, M. Nitta, and M. Ueda, Phys. Rev. Lett. **100**, 180403 (2008).

[83] K. Kasamatsu, H. Takeuchi, M. Nitta, and M. Tsubota, Journal of High Energy Physics **2010**, 68 (2010).

[84] Y. Kawaguchi and M. Ueda, Physics Reports **520**, 253 (2012), spinor Bose–Einstein condensates.

[85] Y. Kawaguchi, M. Kobayashi, M. Nitta, and M. Ueda, Progress of Theoretical Physics Supplement **186**, 455 (2010), http://oup.prod.sis.lan/ptps/article-pdf/doi/10.1143/PTPS.186.455/5304339/186-455.pdf .

[86] E. Merzbacher, American Journal of Physics **30**, 237 (1962), https://doi.org/10.1119/1.1941984 .

[87] G. E. Volovik and V. P. Mineev, Soviet Journal of Experimental and Theoretical Physics Letters **24**, 561 (1976).

[88] D. A. Ivanov, Phys. Rev. Lett. **86**, 268 (2001).

[89] J. Jang, D. G. Ferguson, V. Vakaryuk, R. Budakian, S. B. Chung, P. M. Goldbart, and Y. Maeno, Science **331**, 186 (2011), http://science.sciencemag.org/content/331/6014/186.full.pdf .

[90] K. G. Lagoudakis, T. Ostatnický, A. V. Kavokin, Y. G. Rubo, R. André, and B. Deveaud-Plédran, Science **326**, 974 (2009), http://science.sciencemag.org/content/326/5955/974.full.pdf .

[91] F. Manni, K. G. Lagoudakis, T. C. H. Liew, R. André, V. Savona, and B. Deveaud, Nature Communications **3**, 1309 (2012).

[92] J. Ruostekoski and J. R. Anglin, Phys. Rev. Lett. **91**, 190402 (2003).

[93] S. W. Seo, S. Kang, W. J. Kwon, and Y.-i. Shin, Phys. Rev. Lett. **115**, 015301 (2015).

[94] S.-H. Shinn and U. R. Fischer, Phys. Rev. A **98**, 053602 (2018).

[95] J. Lovegrove, M. O. Borgh, and J. Ruostekoski, Phys. Rev. A **86**, 013613 (2012).

[96] A.-C. Ji, W. M. Liu, J. L. Song, and F. Zhou, Phys. Rev. Lett. **101**, 010402 (2008).

[97] L. M. Symes and P. B. Blakie, Phys. Rev. A **96**, 013602 (2017).

[98] A. Lamacraft, Phys. Rev. B **81**, 184526 (2010).

[99] K. G. Wilson, Rev. Mod. Phys. **47**, 773 (1975).

[100] M. E. Fisher, Rev. Mod. Phys. **46**, 597 (1974).

[101] M. E. Fisher, Rev. Mod. Phys. **70**, 653 (1998).

[102] S. R. White, Phys. Rev. Lett. **69**, 2863 (1992).

[103] S. R. White, Phys. Rev. B **48**, 10345 (1993).

[104] V. Zauner-Stauber, L. Vanderstraeten, M. Fishman, F. Verstraete, and J. Haegeman, Physical Review B **97**, 045145 (2018).

[105] M. M. Wilde, *Quantum Information Theory* (Cambridge University Press, 2013).

[106] A. Pechen, D. Prokhorenko, R. Wu, and H. Rabitz, Journal of Physics A: Mathematical and Theoretical **41**, 045205 (2008).

[107] `https://github.com/AustenLamacraft/tfimps`.

[108] P. Pfeuty, Annals of Physics **57**, 79 (1970).

[109] Wikipedia contributors, "Conjugate gradient method," [Online; accessed 13-June-2019].

[110] S. Mehri, K. Kumar, I. Gulrajani, R. Kumar, S. Jain, J. Sotelo, A. Courville, and Y. Bengio, "Samplernn: An unconditional end-to-end neural audio generation model," (2016), arXiv:1612.07837 .

[111] S. Dieleman, A. van den Oord, and K. Simonyan, in *Advances in Neural Information Processing Systems 31*, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Curran Associates, Inc., 2018) pp. 8000–8010.

[112] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "Wavenet: A generative model for raw audio," (2016), arXiv:1609.03499 .

[113] J. Chung, K. Kastner, L. Dinh, K. Goel, A. C. Courville, and Y. Bengio, in *Advances in Neural Information Processing Systems 28*, edited by C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett (Curran Associates, Inc., 2015) pp. 2980–2988.

[114] N. Kalchbrenner, E. Elsen, K. Simonyan, S. Noury, N. Casagrande, E. Lockhart, F. Stimberg, A. van den Oord, S. Dieleman, and K. Kavukcuoglu, "Efficient neural audio synthesis," (2018), arXiv:1802.08435 .

[115] R. Prenger, R. Valle, and B. Catanzaro, "Waveglow: A flow-based generative network for speech synthesis," (2018), arXiv:1811.00002 .

[116] C. Donahue, J. McAuley, and M. Puckette, "Adversarial audio synthesis," (2018), arXiv:1802.04208 .

[117] J. Engel, C. Resnick, A. Roberts, S. Dieleman, D. Eck, K. Simonyan, and M. Norouzi, "Neural audio synthesis of musical notes with wavenet autoencoders," (2017), arXiv:1704.01279 .

[118] `https://github.com/AustenLamacraft/audio-mps`.

[119] C. Gardiner and P. Zoller, *Quantum World Of Ultra-cold Atoms And Light, Book II: The Physics of Quantum-Optical Devices*, Cold Atoms (Imperial College Press, 2015).

[120] H. M. Wiseman and G. J. Milburn, *Quantum Measurement and Control* (Cambridge University Press, 2009).

[121] T. J. Osborne, J. Eisert, and F. Verstraete, Phys. Rev. Lett. **105**, 260401 (2010).

[122] T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, in *Advances in Neural Information Processing Systems 31*, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Curran Associates, Inc., 2018) pp. 6571–6583.

[123] S. Särkkä and A. Solin, *Applied Stochastic Differential Equations*, Institute of Mathematical Statistics Textbooks (Cambridge University Press, 2019).

[124] P. A. Alvarado and D. Stowell, "Efficient learning of harmonic priors for pitch detection in polyphonic music," (2017), arXiv:1705.07104 .

[125] K. Riley, M. Hobson, and S. Bence, *Mathematical methods for physics and engineering* (Cambridge university press, 1999) p. 1000.

[126] W. J. Wilkinson, M. R. Andersen, J. D. Reiss, D. Stowell, and A. Solin, "Unifying probabilistic models for time-frequency analysis," (2018), arXiv:1811.02489 .

[127] A. Solin and S. Särkkä, in *AISTATS* (2014).

[128] Wikipedia contributors, "Campbell's theorem (probability) — Wikipedia, the free encyclopedia," (2019), [Online; accessed 26-June-2019].